CUKr H05936-21-KP2566

University Library
14.2.07
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

21

VOLUME CIII, NUMBER 7
JULY 2006

page

319 *The Principle of Drift: Biology's First Law* — Robert N. Brandon
336 *Replaying Life's Tape* — John Beatty

COMMENTS AND CRITICISM

363 *Defending Frankfurt's Argument in Deterministic Contexts: A Reply to Palmer.* — Ishtiyaque Haji and Michael McKenna

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2006

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

KP 2566

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.

POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2006 by the Journal of Philosophy, Inc. ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CIII, NO. 7, JULY 2006

University Library CAMPUS

THE PRINCIPLE OF DRIFT: BIOLOGY'S FIRST LAW*

Drift is to evolution as inertia is to Newtonian mechanics. Both are the "natural" or default states of the systems to which they apply. Both are governed by zero-force laws. The zero-force law in biology is stated here for the first time.

The Newtonian analogy to evolutionary theory is fairly common in presentations of population genetics, especially in reference to the Hardy-Weinberg Law. In the philosophy of biology, Elliott Sober's[1] presentation is canonical. Recent literature in the philosophy of biology has both criticized[2] and defended[3] the Newtonian analogy to evolutionary theory. Here I show exactly how the Newtonian analogy is helpful. More interestingly, I show how the analogue of Newton's First Law, what I term The Principle of Drift, is, in a very important way, non-Newtonian. Newtonian or not, it is one of the two fundamental laws of biology (along with the Principle of Natural Selection).

I. NEWTON'S PRINCIPLE OF INERTIA

The focus of this paper is on what is, and what is not, an appropriate zero-force law in Evolutionary Theory. Before turning to that, I want to emphasize a few well-known points about the paradigm of zero-

---

* My thanks to Richard Lewontin, Dan McShea, Fred Nijhout, Grant Ramsey, Mark Rausher, and Alex Rosenberg for helpful comments on earlier drafts of this paper. I also want to thank Greg Wray for help with references and Leonore Fleming for editorial help.

[1] Sober, *Conceptual Issues in Evolutionary Biology* (Cambridge: MIT, 1984, second edition).

[2] See Denis M. Walsh, Tim Lewens, and André Ariew, "Trials of Life: Natural Selection and Random Drift," *Philosophy of Science*, LXIX (2002): 452–73; Mohan Matthen and André Ariew, "Two Ways of Thinking about Fitness and Natural Selection," this JOURNAL, XCIX, 2 (February 2002): 55–83.

[3] See Christopher Stephens, "Selection, Drift, and the 'Forces' of Evolution," *Philosophy of Science*, LXXI (2004): 550–70.

0022-362X/06/0307/319–35

state laws—Newton's First Law, also know as the Principle of Inertia. After reviewing these points I will explore what sort of explanations Newton's First Law supports.

Newton expressed the First Law in two parts, since he thought there was an absolute difference between objects at rest and objects moving at a constant velocity. Of course relativity theory rejects this distinction and therefore we often see simpler one-clause statements of the principle. But for our purposes, the Newtonian version is best. It states that:

(1) An object at rest will stay at rest unless acted on by a net external force.
(2) An object in motion with velocity v will remain in motion with velocity v unless acted on by a net external force.

(Of course, relativity aside, the first clause can be seen as a special case of the second, where $v = 0$.) Newton's First Law represented a major break from the tradition of Aristotelian physics. While Aristotle would have agreed with the first clause stated above, he would have disagreed with the second. For Aristotle the maintenance of constant velocity requires the constant action of a force. (Put anachronistically, for Newton $F = ma$, while for Aristotle $F = mv$.) Thus Newton's First Law changes what calls for an explanation in terms of force. Or to put the point in more Aristotelian terms, Newton fundamentally altered the conception of what constituted "natural" motion.

All this is familiar. But let us now turn our attention to Newtonian explanations of inertial motion. Suppose an object is at rest at time t. Over the time interval t-t' it remains at rest. How do we explain this behavior? The acceleration, a, during this time interval is 0. Applying Newton's Second Law, $F = ma$, we see that $F = 0$. Thus, according to the First Law the object remains at rest. This is elementary physics. We will see if it is elementary philosophy.

What sort of explanation is this? According to the covering-law model of explanation,[4] to explain a phenomenon is to subsume it under some one or more general laws. That is exactly what we did above, and so, based on that model we would say that our explanation is a standard scientific explanation. But if one considers this model of explanation to be inadequate,[5] as I do, then how do we think about this explanation? That, not surprisingly, depends on the alternative model one adopts. It is well beyond the scope of this paper to argue for one theory of explanation over others. But elsewhere I have argued for a causal-mechanical model of explanation in evolutionary biology.[6] Since our ultimate concern is with evolutionary biology, let us ask the question of how that model treats the above bit of Newtonian physics.

---

[4] See, for example, Carl Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965).

[5] See Wesley Salmon, *Statistical Explanation and Statistical Relevance* (Pittsburgh: University Press, 1971), and his *Scientific Explanation and the Causal Structure of the World* (Princeton: University Press, 1984).

If one thinks of Newtonian forces as Newtonian causes—and that certainly seems reasonable—then one might think that the explanation of inertial phenomena is noncausal. Our explanation above did not cite a force; therefore it did not cite a cause. Rather it cited the absence of causes. So how can it be a causal explanation? One response to this question seems so unattractive that I will simply mention it without laboring to refute it. One might say that Newton's First Law tells us what does, and what does not, require explanation. Inertial phenomena do not. Although there is something right about this response, and we will turn to that shortly, saying a whole class of well understood phenomena cannot be explained seems just wrong. A second response is to say that the above reasoning is correct, that the explanation is noncausal, and that shows the limitations of the causal-mechanical model of explanation. Although a number of philosophers of science would be sympathetic to this response, I am not. Salmon (*op. cit.*) argues that the goal of explanatory science is to uncover the causal structure of the world. The Principle of Inertia (or rather its Relativistic counterpart) seems to be a fundamental part of the causal structure of our world. Because of that, I would consider our inertial explanation to be a perfectly good causal-mechanical explanation.

But, returning to the point made above, there is something different about this sort of Newtonian explanation versus one that explains some noninertial phenomenon in terms of some net force acting on the object. The latter sort of phenomenon requires "special" explanations. That is the lesson of Newton's First Law. Thus we could mark this distinction by saying that noninertial phenomena require "special-causal-explanations," while inertial phenomena require what I will call "default-causal-explanations."

In Newtonian mechanics, the Principle of Inertia clearly delimits what can, and cannot, be explained by default explanations. That is, it delimits the class of objects that fall under the zero-force law. We will see if there is any analogous law in biology.

---

[6] See Brandon, *Adaptation and Environment* (Princeton: University Press, 1990), where I was inspired by Salmon's *Scientific Explanation and the Causal Structure of the World.*

II. THE HARDY-WEINBERG LAW AS A ZERO-FORCE LAW

The existence of a zero-force law presumes the existence of forces. Are there forces of evolution? John A. Endler[7] argued that the analogy to Newtonian physics was more misleading than helpful. More recently, Dennis Walsh, Tim Lewens, and André Ariew (*op. cit.*), and Mohan Matthen and Ariew separately (*op. cit.*) have argued against the force analogy. I will not try to respond to these arguments in a point by point fashion; in part because I have offered detailed criticisms of the "emergentist statistical interpretation" of evolutionary theory defended in the two last mentioned papers,[8] but also because I think a direct positive case can be made for considering certain evolutionary processes as forces.

Evolutionary Forces. The processes I have in mind are selection, mutation, migration, and nonrandom mating. In physics a force is a vector quantity. That means it has both a magnitude and a direction. Also various forces in Newtonian physics (for example, gravity, electrical force, friction) can all be measured in a standard quantitative unit (à la Newton). Is there a common currency to measure the processes mentioned above? From a population genetics point of view all can be measured in terms of their effects on gene frequencies (see caveat below). From the point of view of molecular evolution they would be measured on a finer grain—the level of gene sequences. And one could take a purely phenotypic point of view and measure in terms of phenotypic distributions. Let us ignore this embarrassment of riches and focus on the population genetics perspective.

Consider mutation and migration. Both are similar in that they are already net forces that could be broken down into their components. Migration consists of immigration into a population and emigration from it. Both can have effects on gene frequencies, which can then be added together to get the force of migration; similarly for mutation. In a simple system with one locus and two alleles, A_1 and A_2, there is an actual number of mutations from A_1 to A_2, and an actual number from A_2 to A_1. The change in A_1 is simply the second number minus the first. There is no problem in conceiving of these as forces. (Although some might insist on considering them as net forces rather than as component forces: that would not militate against the view that evolutionary theory is a theory of forces.)

---

[7] Endler, *Natural Selection in the Wild* (Princeton: University Press, 1986).

[8] See Brandon and Grant Ramsey, "What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift," in Michael Ruse and David L. Hull, eds., *The Cambridge Companion to Philosophy of Biology* (New York: Cambridge, forthcoming).

Nonrandom mating is different. I mention it now, because we are about to consider the Hardy-Weinberg Law, and nonrandom mating is one of the processes that can cause a population to deviate from H-W equilibrium. Nonrandom mating has two basic forms: positive assortative mating and negative assortative mating. Let me say explicitly that here we are solely concerned with forms of nonrandom mating that do not involve differential reproduction among different genotypes. (So, for example, we are not here talking about positive assortative mating where rare genotypes have lower reproductive success because of difficulty in finding mates.) Nonrandom mating has an immediate effect on genotype frequencies, not gene frequencies. (Consider a population with two alleles, A_1 and A_2, at frequencies p and q. If the homozygotes preferentially mate with each other, p and q will remain unchanged but there will be a decrease in the number of heterozygotes.) Nonrandom mating can have selective effects (for example, sexual selection or selection against inbreeding), but from the point of view of the H-W law it is its effect on genotypic frequencies that matters. This contradicts the statement made above that all of these processes can be measured in terms of effects on gene frequencies. But we can achieve consistency by measuring mutation and migration in terms of effects on genotypic frequencies. From a population genetic point of view that is inconvenient, but possible. So I consider this a minor technical point and will henceforth ignore it.

Selection presents a new challenge to the view that these evolutionary processes are forces. People tend to think of directional selection when they think of selection (this point will be explored and defended below). The other two patterns of selection are stabilizing and disruptive. These terms come from quantitative genetics so are usually defined in terms of phenotypic distributions. Thus directional selection is when one extreme in the distribution is favored (for example, the taller the fitter). Stabilizing selection is when a single point in the distribution is favored (for example, 5m tall is the fittest). Disruptive selection occurs when 2 (or more) points in the distribution are favored (for example, 2m tall and 6m tall are the fittest). But these terms have come to be used more generally and can be translated into genetic terms. For instance in a two-allele model, if A_1 is rare and A_1A_1 is fitter than A_1A_2, which is fitter than A_2A_2, then we would call that directional selection. But remember that a force is supposed to have a magnitude and a direction. This seems totally unproblematic for directional selection. But how about the other two forms? What is the direction of stabilizing selection? If a population, or some subpart of it, finds itself to the left of the selected point, then the direction of the force is to the right. But if the population is on the right of that point then the direction is to the left. Is our force analogy in trouble? No, or rather, only if it is problematic to think of gravity as a force. In our solar system two objects of equal mass and equal distance from the Sun, but on opposite sides of the Sun, will experience a force of equal magnitude but opposite direction. That is exactly analogous to two subpopulations equidistant from the selected point in stabilizing selection but on opposite sides (assuming the selection gradient is the same on both sides). That gravitational force is relative to position is obvious in Newtonian mechanics. So that a similar relativity exists with respect to stabilizing and disruptive selection is not problematic.

Thus, although it has not been exactly straightforward, we have made a positive case for treating certain evolutionary processes as forces. Notice that drift has yet to be mentioned.

The Hardy-Weinberg Law. There are two importantly different ways of stating the H-W law. One statement of it lumps drift together with mutation, migration, and the rest. The other does not. Here are the two versions:

> H-W$_1$: If a population exists with two alleles, A_1 and A_2, with frequencies p and q respectively, then in a single generation the population will settle into genic and genotypic equilibrium with gene frequencies p and q, and genotypic frequencies of $A_1A_1 = p^2$; $A_1A_2 = 2pq$; and $A_2A_2 = q^2$—provided that there is no selection, mutation, migration, nonrandom mating, or *drift.*
>
> H-W$_2$: If an *infinite* population exists with two alleles, A_1 and A_2, with frequencies p and q respectively, then in a single generation the population will settle into genic and genotypic equilibrium with gene frequencies p and q, and genotypic frequencies of $A_1A_1 = p^2$; $A_1A_2 = 2pq$; and $A_2A_2 = q^2$—provided that there is no selection, mutation, migration, or nonrandom mating.

If H-W$_1$ is to be considered a zero-force law then drift needs to be a force just like the other evolutionary processes. However, I think that there are overwhelming reasons for not considering it a force.

Consider first the idea that a force has both a magnitude and direction. Drift has a magnitude that can be probabilistically predicted prior to the fact, and can be quantitatively accessed after the fact.[9] But drift definitely does not have a direction. (That, of course, is why it is called *drift.*) Given a population consisting of two selectively neutral alleles, A_1 and A_2, at frequencies $p = q = .5$, we can predict that one of the two alleles will go to fixation (in the absence of evolutionary forces), but we cannot predict which will. That means drift has no direction. Christopher Stevens (*op. cit.*) argues that drift does have a direction at the genotypic level. He suggests that it predictably leads to loss of heterozygosity and increase of homozygosity. That is a prediction, but a prediction without a direction. It does not say which of A_1A_1 or A_2A_2 will increase in frequency. In physics that would be like saying that a 20-Newton force is acting on object *A*. Such a statement either makes no sense (the magnitude, but not direction has been specified) or is incomplete (oops, I meant a 20-Newton downward force). Notice that we were able to specify the direction of the other evolutionary processes mentioned in H-W_1.

---

[9] See Brandon, "The Difference between Selection and Drift: A Reply to Millstein," *Biology and Philosophy*, xx (2005): 153–70.

Second, and perhaps more importantly, drift is not a "special" force in evolution; it is the default position. By that I mean that it is part and parcel of a constitutive process of any evolutionary system—namely, the sampling process. Sampling takes place in a number of ways. I will not try to exhaustively name all of them, because I cannot, and I suspect no one can. That is, I suspect there are as of yet undiscovered sampling processes going on in evolutionary systems. But here is a partial list of sampling processes. Evolution requires reproducing entities that form lineages, parent-offspring lineages. So in many cases there is sampling of parents (not all potential parents actually reproduce). In sexual diploid organisms there is gametic sampling in the formation of a new diploid cell (fertilized zygote). There is also chromosomal sampling in the formation of gametes. There is further sampling of parts of chromosomes in crossover events during meiosis. At the ecological level there is habitat sampling when organisms are dispersed over heterogeneous selective environments.[10] The process that leads to drift, sampling, is a necessary part of any evolutionary system. It is not some new process added to the basic evolutionary process. In contrast, mutation, migration and nonrandom mating are clearly separable processes, which may, or may not, be a part of some particular evolutionary scenario.

(Asking if drift is a force is just like asking if inertia is a force in Newtonian mechanics. The lesson of Newton's First law is that it is not.)

Selection, I have argued[11] is both conceptually and empirically distinguishable from drift. When nonequal probabilities govern the sampled entities, selection becomes possible. However the possibility of drift does not go away just because the equiprobable distribution is no longer in force. Drift becomes impossible only in the highly implausible condition that I have labeled a state of *Maximal Probability Difference*—that is, when all of the fitnesses of the competing entities equal either 1 or 0, with both values being present. (See Figure 1.)

---

[10] See Brandon, *Adaptation and Environment*, especially chapter 2, "The Concept of Environment in the Theory of Natural Selection."

[11] See Brandon, "The Difference between Selection and Drift."

Selection, like drift, depends on the sampling process. However, unlike drift, it has a direction and is separable from drift. Thus I think it is reasonable to treat selection as a "special" evolutionary force (on a par with mutation, and so on), while it is clearly unreasonable to think of drift that way. I will offer more support for this position shortly. For now, let us tentatively adopt the position that drift is not a force. It follows that H-W_1 is not a zero-force law. For it to be true it needs to mix legitimate evolutionary forces with one non-force (drift). H-W_1 is useful in many ways. It is true. It just is not a zero-force law.

H-W_2 does not run afoul of the above problem. It mentions only legitimate evolutionary forces in its proviso-clause. However, it is

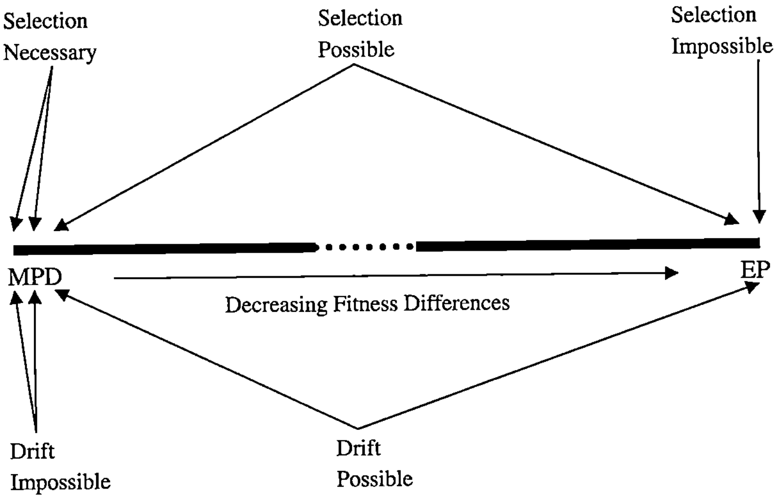


Figure 1: The heavy horizontal line, with dotted center section, represents the infinite number of possible fitness distributions from Maximal Probability Differences (MPD—all fitness = 0 or = 1, with some of both) on the left to the Equiprobable Distribution (EP—all fitness the same) on the right. The arrows emanating from the different descriptions of the modalities of selection and drift indicate the areas of the distribution falling under these descriptions.

unclear how it applies to real populations, all of which are finite. I am not criticizing H-W_2 on the basis of its employing an idealization. That is normal and useful in science. But when we talk about frictionless planes in physics and get results that apply in that idealized situation, we ultimately want to be able to apply it to real world situations. So how do we apply H-W_2 to finite populations? One answer is that we do not. I presume that most would be unsatisfied with this. A second answer is that in applying H-W_2 to finite populations we transform it back into H-W_1; that is, we stick drift back into the proviso-clause. But then, as we have seen, we no longer have a zero-force law.

A final answer is that we do not revert to H-W_1, we leave the provisos alone, and we assert H-W_2 is true of finite populations. Then H-W_2 basically says of finite populations: If no evolutionary forces act on a population then its gene and genotypic frequencies remain unchanged. That is logically equivalent to the following: If a population changes in gene or genotypic frequencies then some evolutionary force is acting on it. Assuming, as we are for the moment, that I am right in not classifying drift as an evolutionary force, then these statements are false. A population with no evolutionary forces acting on it will drift. (And so, a population that changes over generational time may or may not be experiencing evolutionary forces.) This truth is exactly what guides contemporary work on trying to distinguish selection from drift. For instance, at the molecular level we compare synonymous substitutions (substitutions in the third place in a codon that end up coding for the same amino acid), with nonsynonymous substitutions to see where, and with what strength, selection is acting.[12]

H-W_2 applied in a straightforward manner to finite populations is not only false in what it says, it is also misleading in what it suggests. The first gloss on H-W_2, stated in the paragraph above, basically says: If no forces then no change. That does not strictly imply the following: If no change then no force; but it does suggest it. It suggests that lack of change is a sign of lack of force. (This suggestion is clear in Sober's Newtonian analogy (*op. cit.*).) But in fact, the opposite is true: If no change then a set of forces must be acting. Again this truth is important in that it licenses the sorts of inferences biologists need to make in understanding equilibrium situations (more on this shortly).

---

[12] See, for example, Martin Kreitman, "Methods to Detect Selection in Populations with Applications to the Human," *Annual Review of Genomics and Human Genetics*, I (2002): 539–59; Ziheng Yang and Joseph P. Bielawski, "Statistical Methods for Detecting Molecular Adaptation," *Trends in Ecology and Evolution*, XV (2000): 496–503; Michael Bamshad and Stephen P. Wooding, "Signatures of Natural Selection in the Human Genome," *Nature Reviews Genetics*, IV (2003): 99–111.

Thus H-W_2 is either: (a) A true zero-force law inapplicable to any real population; or (b) Applicable, but then equivalent to H-W_1, and so not a zero-force law; or (c) Applicable, but false.

And so the Hardy-Weinberg Law fails to be a zero-force law in evolutionary biology. Is there a zero-force law in evolutionary theory?

III. THE PRINCIPLE OF DRIFT AS A ZERO-FORCE LAW

The answer is "yes." But the law in question has heretofore been unstated. I will state it first, and then defend its truth, its usefulness, and its status as a zero-force law.

The Evolutionary Principle of Drift:

> (A) A population at equilibrium will tend to drift from that equilibrium unless acted on by an evolutionary force.
> (A population at rest will tend to start moving unless acted on by an external force.)
>
> (B) A population on evolutionary trajectory *t*, caused by some net evolutionary force F, will tend to depart from the extrapolated path predicted based on F alone (in either direction or magnitude or both) even if no other evolutionary force intervenes, unless F continues to act.
> (A population in motion will tend to stay in motion, but change its trajectory, unless continually acted on by an external force.)

The Principle of Drift (PD) is stated in two clauses, just like Newton's version of his Principle of Inertia. I have stated each clause first in more precise biological language and then, in parentheses, in a more physics-like language that facilitates a comparison to Newton's First Law. Clause (B) is clearly non-Newtonian. In fact, it is Aristotelian in that it requires the constant action of a force to keep a population in the same state of motion. But Aristotle would not be happy with Clause (A), which is both non-Newtonian and non-Aristotelian. So in this way the PD is quite unlike Newton's First Law.

However, in another way, there is a strict parallel to Newton's First Law. Recall our discussion above where we saw how Newton's law sets out inertia as the default state of an object. I argued that Newtonian explanations of inertial states are importantly different from Newtonian explanations of non-inertial states. The latter require the citation of "special" Newtonian causes, while the former are what I called default-causal-explanations. The PD serves exactly the same purpose. It clearly sets out the default state of a population, or in other words, it is a principle of default evolutionary action. In the absence of evolutionary forces drift occurs, or, more precisely drift will tend to occur (see Figure 1 above). Do not misunderstand this, drift can certainly occur in the presence of evolutionary forces as well. So, for instance, in clause (B) above, a population continually acted on by net evolutionary force, F, will quite frequently depart from the trajectory predicted from F alone. F needs to be very strong relative to the effective population size for it to dominate drift.[13] Drift is the default state of evolutionary systems.

We can explain evolutionary change due to drift. The drift producing potential of the sampling processes that are constitutive of the evolutionary process is a fundamental part of the causal structure of our world. Thus such explanations are causal but differ from evolutionary explanations that cite specific evolutionary forces. Just as in the Newtonian case, the first sort of explanation is a default-causal-explanation and the second a special-causal-explanation. It is the PD that enables us to make this distinction.

Is the PD true? I think its truth is patently obvious once one potential misunderstanding of it is removed. One might think that the PD takes a stand on an empirical issue, the truth of which is in doubt. That is, one might think that the PD is somehow committed to some particular claim about the relative importance of drift versus selection. But this is a misunderstanding. Newton's First Law says nothing about the relative frequency of inertial objects versus objects experiencing a net force in the universe. Indeed, think of the paradigm example of a system to which Newton's Laws apply—our solar system. All of the major objects in that system are experiencing constant acceleration due to gravitational forces. Likewise, the PD is perfectly compatible with the claim that selection and other evolutionary forces are the primary movers of the vast majority of populations.

Consider clause (A). It says that in the absence of evolutionary forces (and in particular, selection) a population previously at equilibrium will start to drift. The empirical debate over the relative importance of drift versus selection is irrelevant to this claim. By hypothesis clause (A) removes selection. Then, by hypothesis, the existing variants are selectively neutral. From that it follows that drift will occur.

---

[13] See Brandon and H. Fred Nijhout, "The Empirical Non-equivalence of Genic and Genotypic Models of Selection: A (Decisive) Refutation of Genic Selectionism and Pluralistic Genic Selectionism," forthcoming in *Philosophy of Science.*

Consider clause (B). It says that in the absence of evolutionary forces (and in particular, selection) a population previously moving in state space with trajectory *t* will depart from that trajectory. Again, the truth of this is fairly obvious. If the trajectory were due to one or more of the special evolutionary forces, then removing those forces would tend to change the trajectory.

So the truth of the Principle of Drift is not hostage to any particular resolution of the empirical selectionist/neutralist debate. However, some philosophers might think its truth is dependent on a questionable metaphysical assumption—namely, that the evolutionary process is fundamentally indeterministic. First, the question of determinism versus indeterminism for the whole world, or for some part of it (the evolutionary process), is an empirical question, not a metaphysical question. I believe, and have defended the view, that the evolutionary process is indeed indeterministic,[14] but I think that the only assumption that has been at work in describing the Principle of Drift is that drift is a real phenomenon. Drift happens. Drift is distinguishable from other sorts of evolutionary change. No biologists doubt this. And that is all I have assumed. The (misguided) metaphysical determinist needs to explain how drift can exist in a fundamentally deterministic world. (Since that is a fantasy world I leave that exercise to him or her.)

It is also a misunderstanding to think that I have defended the Principle of Drift at the expense of the H-W Law. I have argued that the H-W Law is not a zero-force law, but I have not denied its utility. In fact its strength relative to the Principle of Drift flows from its limitations. The H-W Law is explicitly couched in the terms of the population genetics of diploid sexual organisms. That allows it, for instance, to be used to show that assortative mating is occurring in a population when genotypic frequencies depart from the H-W prediction. The PD is not couched in this specific language and could not be used for this particular inference. But, as John Beatty[15] pointed out some time ago, this very specificity shows that the H-W Law is no law at all. It depends on conditions that are evolutionarily derived, and not at all necessary for the evolutionary process. The evolutionary forces that brought the H-W into play could change and render H-W totally inapplicable to life on Earth (for example, if all lineages were to go asexual). The Principle of Drift is different. It is not predicated on diploidy or sex. It is based solely on the fact that evolution consists of one or more sampling processes. In that way it is perfectly general and is a law of nature.

---

[14] See Brandon and Scott Carson, "The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables Proof' but No Room for Determinism Either," *Philosophy of Science*, LXIII (1996): 315–37.

[15] Beatty, "What's Wrong with the Received View of Evolutionary Theory?" in P.D. Asquith and R.N. Giere, eds., *PSA 1980*, Volume 2 (East Lansing, MI: Philosophy of Science Association, 1980), pp. 397–426.

One way to think about the lawfulness of biological generalizations is to ask if they would be exportable to life elsewhere in the universe. Were I asked to take Evolutionary Theory to some sentient beings elsewhere in the universe I would pack the Principle of Natural Selection. Where there is life there is natural selection. I would also pack the Principle of Drift. Along with a few personal items I would also pack my math books. And that is all. I would not bother with the H-W Law because I would not necessarily expect to find diploid sexual organisms. If I did find them I would realize that $(p+q)^2 = p^2 + 2pq + q^2$ and would thereby recreate H-W.

So although my point is not to criticize the H-W Law, it is clear that it is not fundamental, unlike the PD. Likewise, and perhaps more importantly, it does not provide appropriate null hypotheses in evolution. It suggests evolutionary stasis as a null hypothesis. But this is just wrong. In the absence of evolutionary forces drift occurs. So drift is the appropriate null hypothesis.

Finally, and relatedly, the H-W law is not a zero-force law. The PD is. It is in virtue of this fact that it does provide the appropriate null hypothesis.

IV. CONSEQUENCES

Exactly how Evolutionary Theory relates to Newtonian mechanics may seem an esoteric topic in philosophy of biology. It is not. It has real consequences. I have explored the biological consequences of the PD elsewhere.[16] Here the focus will be on the more purely philosophical.

First, the very concept of evolution suggests the view that stasis is the default condition of evolutionary systems and that change is what is fundamental to evolution. Thus evolution is often defined as *change* in gene frequencies. Those objecting to a narrow population genetics perspective have suggested other definitions, such as, evolution is *change* in phenotypic distributions, or evolution is *change* in developmental programs. But from the perspective of this paper, all such definitions are entirely misleading. Change is a heterogeneous category in evolution. It may be driven by special evolutionary forces, or it may result from no net forces at all. Stasis, on the other hand, seems to be quite homogeneous in its cause. There is no way to produce long-term stasis other than strong stabilizing selection.

---

[16] See Brandon and Nijhout; and see D. McShea and Brandon, "Diversity and Complexity: the Zero-Force Expectation for Biology" (in preparation).

Second, this privileging of change, the failure to see the fundamentally different causes of change (default versus special), and the failure to see stasis as a strong signal of selection, together have led to empirically inadequate and explanatorily empty conceptions of fitness. Thus, starting with R.A. Fisher,[17] many have sought to *define* fitness in terms of evolutionary *change*. (For Fisher, fitness is the per capita rate of increase of a type.) Recently Matthen and Ariew (*op. cit.*) have argued that evolutionary biologists *must* measure fitness in terms of evolutionary change. But if one of the primary ways selection acts is in stabilizing selection, then these views are entirely wrong-headed. Strong stabilizing selection (or what is called *purifying selection* at the molecular level) results in no change (in gene frequencies, or at the molecular level, in DNA sequence). And indeed, at the molecular level the strongest evidence of the past action of selection is evolutionary conservation of DNA sequence. (As, for example, in the PAX6 gene found in virtually identical forms in fruit fly, zebrafish, mice, and humans.[18]) But the Fisherian is conceptually precluded from seeing this as strong selection, since nothing is changing. Further, it should be noted against the argument that biologists must/should measure fitness in terms of evolutionary change, that in fact, biologists measure fitness in that way only very rarely and only as a last resort.[19]

Finally, the views put forward in this paper have implications for what is perhaps the most widely debated question in philosophy of biology—namely, whether or not there are laws in biology. I used to think that biology had exactly one law—the Principle of Natural Selection.[20] I now see it has two.

Perhaps more interesting than counting the number of fundamental laws in evolutionary theory is recognizing the thoroughly probabilistic nature of those laws. I have argued extensively elsewhere[21] that the Principle of Natural Selection is an instantiation of (a specially structured version of) what is called the Principle of Direct Inference in Probability Theory.[22] This is the principle that allows one to infer from statistical probabilities (or chances) to actual frequencies.[23] For example, if coin A has a probability of heads = 0.7 (in a certain chance set-up) while coin B has a probability of heads = 0.3 (in the same chance set-up), then if we toss both coins 10 times (in the same set-up) then (probably) coin A will land on heads more than coin B. That is an application of the Principle of Direct Inference. Consider the PNS: If organism A is fitter than organism B in environment E, then (probably) A will have more offspring in E than B. Clearly, this is a perfectly analogous application of the same principle. In this paper we have seen how the Principle of Drift is purely the result of the finite sampling processes that are part and parcel of the evolutionary process. Thus, for those who are reductionistically minded (I am not), these two laws point to probability theory as the reductive foundation of evolution, not to physics or chemistry.

---

[17] Fisher, *The Genetical Theory of Natural Selection* (New York: Oxford, 1930).

[18] See, for example, R. Quiring, U. Walldorf, U. Kloter, and W.J. Gehring, "Homology of the Eyeless Gene of *Drosophila* to the *Small Eye* Gene in Mice and *Aniridia* in Humans," *Science*, CCLXV (1994): 785–89.

[19] See Endler; Brandon and Ramsey; and J.G. Kingsolver, H.E. Hoekstra, J.M. Hoekstra, D. Berrigan, S.N. Vignieri, C.E. Hill, A. Hoang, P. Gibert, and P. Beerli, "The Strength of Phenotypic Selection in Natural Populations," *The American Naturalist*, CLVII (2001): 245–61.

[20] See Brandon and Alex Rosenberg, "Philosophy of Biology," in Peter Clark and Katherine Hawley, eds., *Philosophy of Science Today* (New York: Oxford, 2003), pp. 147–80.

---

[21] See Brandon, "Adaptation and Evolutionary Theory," *Studies in History and Philosophy of Science*, IX (1978): 181–206, and *Adaptation and Environment*.

[22] See Ian Hacking, *Logic of Statistical Inference* (New York: Cambridge, 1965), pp. 1–12.

[23] There are other interpretations of this principle, which instead of connecting objective probabilities with outcomes, connect them with subjective probabilities; see Isaac Levi, "Direct Inference," this JOURNAL, LXXIV, 1 (January 1977): 5–29. My view is that an objectivist theory of probability requires principles to connect observation of relative frequencies to objective probabilities and to connect objective probabilities with observable events. On this view probabilities are no different from any other sort of theoretical entity, for example, black holes. If we are meaningfully to posit black holes we need principles to go from observables to black holes and vice versa; similarly for objective probabilities. Subjective probabilities are superfluous on this view. Given these principles we can reconstruct subjective probabilities, but they do no work. My view of evolutionary theory's requirement of objective probabilities is similar to Karl Popper's view of quantum mechanics. See Popper, "The Propensity Interpretation of Probability," *The British Journal for the Philosophy of Science*, X (1959): 25–42. My view of objective probability finds precedent in the works of Hans Reichenbach, Salmon, and Hacking, among others. In particular, I follow Salmon's idea of objectively homogeneous reference classes as being the solution to the reference class problem. I have specifically developed these ideas in the context of evolutionary biology and have argued that my notion of "selective environmental homogeneity" solves the reference class problem in evolutionary biology. For the most recent development of this argument, see Brandon, "The Difference between Selection and Drift: A Reply to Millstein," *Biology and Philosophy*, XX (2005): 153–70. These issues involving the interpretation of probability and the principle of direct inference are enormously complex, important, and well beyond the scope of this paper. I would only add that my view of them seems to be just what evolutionary theory requires.

V. BIOLOGY'S FIRST LAW

Newton's First Law is so called not because it was the first law of physics discovered or articulated. Here "First" does not relate to temporal priority. Rather it describes a sort of logical, or better, nomological priority. Newton's Law of Inertia describes the default, or "natural," state of Newtonian systems. It describes what happens when no net Newtonian forces are acting on such a system. It is, as we have seen, a zero-force law. I have shown that The Principle of Drift is analogous to Newton's First Law in these ways. That is why I call it biology's first law.[24]

But from a temporal point of view the Principle of Drift is new, or at least relatively new. However, in the paper that introduced the Hardy-Weinberg law, G.H. Hardy talks about a drift phenomenon, without using that word.[25] This is probably the first discussion of evolutionary drift.[26] It would be an interesting historical project to see just how and when drift as the default state of evolutionary systems became embedded into evolutionary thinking. My point is that, unlike the Principle of Natural Selection, which one might reasonably trace back to Darwin, the Principle of Drift is relatively modern. It has been first named and articulated in this paper. However, contemporary evolutionary biologists regularly make the inferences it licenses. So I do not claim to have made a new biological discovery here. What I have done is to systematize these important inferences into a single law, a law that is fundamental to contemporary evolutionary biology.

VI. CONCLUSIONS

How similar is modern evolutionary theory to Newtonian mechanics? Taken in isolation, that question is at best silly, at worst meaningless. Is the answer 17? Taken in the context of recent literature in the philosophy of biology we can say that it is not as similar to Newtonian mechanics as Christopher Stevens thinks, but much more so than either Walsh, Lewens, and Ariew, or Matthen and Ariew would allow. What I have found surprising is how useful is the comparison. Newton's First Law provides the model for a zero-force law. The Principle of Drift, which is stated here for the first time, is in one important way non-Newtonian. For a population to maintain its trajectory through state-space the constant action of evolutionary forces is required. But, the Principle of Drift is like Newton's Principle of Inertia in that it is a zero-force law, and so it does provide appropriate null-hypotheses. The Hardy-Weinberg law fails to be a zero-force law. It fails not just because it is not a law, but also because it mixes evolutionary forces (selection, migration, mutation, nonrandom mating) with a nonforce (drift). And thus it fails to provide the appropriate null-hypothesis for evolutionary scenarios.

---

[24] In calling this biology's first law I do not mean to suggest that evolutionary biology is the whole of biology. However, I do mean to suggest something that virtually all biologists would agree with, namely, that evolutionary biology lies at the center of biology.

[25] G.H. Hardy, "Letter to the Editor," *Science*, XXVIII (1908): 49–50.

[26] Richard Lewontin, personal communication, 2005.

ROBERT N. BRANDON

Duke University

REPLAYING LIFE'S TAPE*

It might seem self-evident that evolutionary biology is an "historical" discipline. But in the mid-to-late nineteen seventies and early eighties, a number of evolutionary biologists began to argue that the field was lacking historical perspective.[1] Among the most outspoken proponents of this view was Stephen Gould, for whom taking history seriously meant—above all—taking "contingency" seriously. As he reflected in his book *Wonderful Life*, "...the essence of history. Its name is contingency...."[2]

To explain what he meant, Gould offered a thought experiment:

> I call this experiment "replaying life's tape." You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past.... Then let the tape run again and see if the repetition looks at all like the original (WL 48).

---

*This paper is dedicated to the memory of Steve Gould, who was so thoughtful, constructive, and generous in his engagements with historians and philosophers of science. The paper was originally prepared for a special session in his honor at the 2002 Philosophy of Science Association meetings. I am grateful to the PSA Program Committee for the invitation. I am also grateful to Jonathan Losos, Richard Lenski, Michael Travisano, Michael Whitlock, Sally Otto, Sandra Mitchell, Philip Kitcher, Elliott Sober, Marc Ereshevsky, Maggie Osler, Roberta Millstein, Sahotra Sarkar, Scott Anderson, Catherine Wilson, Alan Richardson, Chris Stevens, Mohan Matthen, and Ori Simchen, for comments and criticisms.

[1] There had been earlier attempts to stress the importance of "history" in evolutionary biology, for example, R.C. Lewontin, "The Principle of Historicity in Evolution," in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution* (Philadelphia: Wistar Institute, 1967). But something like a "movement" developed in the nineteen-seventies and eighties. There were also, around this time, calls for greater attention to history in systematics, biogeography, ecology, and conservation biology (for an insightful overview, see S.E. Kingsland, "Afterword," in *Modeling Nature: Episodes in the History of Population Ecology* (Chicago: University Press, 1995), pp. 213–51). There were calls for taking history more seriously in the social sciences as well. See, for example, J.H. Goldthorpe, "The Uses of History in Sociology: Reflections on Some Recent Tendencies," *British Journal of Sociology*, XLII (1991): 211–30; and J.M. Bryant, "Evidence and Explanation in History and Sociology: Critical Reflections on Goldthorpe's Critique of Historical Sociology," *British Journal of Sociology*, LXV (1994): 3–19. It would be very interesting to investigate the various concepts of "history" involved, and the reasons why history became so important in so many areas at this time. My aims here are much narrower.

[2] Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (New York: Norton, 1989), p. 51 (hereafter in citations WL).

0022-362X/06/0307/336–62

His expectation was that, "any replay of the tape would lead evolution down a pathway radically different from the road actually taken" (WL 51).

In this paper, I will examine Gould's thesis in connection with two interestingly different attempts to test it: (1) Jonathan Losos and colleagues, "Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards," and (2) Michael Travisano and colleagues, "Experimental Tests of the Roles of Adaptation, Chance, and History in Evolution."[3] The first is based on a macroevolutionary, "natural experiment." The second is based on a laboratory-controlled, microevolutionary experiment. There many other interesting studies that bear on Gould's views. I have chosen to focus on these two because they explicitly address Gould's claims, and because they are very often cited in this regard.[4]

---

[3] J.B. Losos, T.R. Jackman, A. Larson, K. de Queiroz, and L. Rodríguez-Schettino, "Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards," *Science*, CCLXXIX (1998): 2115–18; M. Travisano, J.A. Mongold, A.F. Bennett, and R.E. Lenski, "Experimental Tests of the Roles of Adaptation, Chance, and History in Evolution," *Science*, CCLXVII (1995): 87–90.

[4] For references to the Losos and Travisano studies in connection with Gould's views, see S.B. Emerson, "A Macroevolutionary Study of Historical Contingency in the Fanged Frogs of Southeast Asia," *Biological Journal of the Linnean Society*, LXXIII (2001): 139–51; P.H. Harvey and L. Partridge, "Different Routes to Similar Ends," *Nature*, CCXCII (1998): 552–53; R.B. Huey, G.W. Gilchrist, M.L. Carlson, D. Berrigan, and L. Serra, "Rapid Evolution of a Geographic Cline in Size in an Introduced Fly," *Science*, CCLXXXVII (2000): 308–09; R.C. MacLean and G. Bell, "Divergent Evolution during an Experimental Adaptive Radiation," *Proceedings of the Royal Society of London B*, CCLXX (2003): 1645–50; M. Matos, T. Avelar, and M.R. Rose, "Variation in the Rate of Convergent Evolution: Adaptation to a Laboratory Environment in *Drosophila subobscura*," *Journal of Evolutionary Biology*, XV (2002): 673–82; E.B. Taylor and J.D. McPhail, "Historical Contingency and Ecological Determinism Interact to Prime Speciation in Sticklebacks, *Gasterosteus*," *Proceedings of the Royal Society of London B*, CCLXVII (2000): 2375–84; and B. Vanhooydonck and D.J. Irschick, "Is Evolution Predictable? Evolutionary Relationships of Divergence in Ecology, Performance and Morphology in Old and New World Lizard Radiations," in P. Aerts, A. D'Août and R. Van Damme, eds., *Topics in Functional and Ecological Vertebrate Morphology* (Maastricht: Shaker, 2002), pp. 191–204.

For an interesting contrast with the natural and laboratory experiments reported by the Losos and Travisano studies, see the artificial life experiments of G. Yedid and G. Bell, "Macroevolution Simulated with Autonomously Replicating Computer Programs," *Nature*, CDXX (2002): 810–12; and D. Wagenaar and C. Adami, "Influence of Chance, History and Adaptation on Evolution in *Digitalia*," *Artificial Life*, X (2004): 181–90. Both explicitly address Gould's views on contingency.

Another major research program that has clear bearing on Gould's views—but that is not explicitly cast in those terms—is Dolph Schluter and his colleagues' investigations of parallel, seemingly replayable, adaptive radiations in threespine sticklebacks: D. Schluter and J.D. McPhail, "Ecological Character Displacement and Speciation in Sticklebacks," *American Naturalist*, CXL (1992): 85–108; Schluter, "Adaptive Radiation in

Another reason for discussing these two studies is that they are based on different interpretations of Gould's views. And yet, as I will argue, both interpretations are fair. That is, Gould—seemingly inadvertently—articulated two quite different notions of historical contingency.

To begin, I will briefly introduce and contrast the two versions of contingency that Gould articulates in *Wonderful Life*. Then I will discuss the two versions in detail in order to show more clearly how they differ, and also in order to show that they can be seen as *complementary* rather than competing accounts of contingency. For example, they play complementary roles in Gould's critique of the all-sufficiency of selectionist reasoning in evolutionary biology.

Next, I will discuss the Losos and Travisano studies, and how they bear upon the two versions of contingency.

Finally, in my conclusion, I will briefly suggest an additional way of reconciling Gould's two versions of contingency (that is, in addition to stressing the complementary roles that they play in his critique of selectionist reasoning), namely, by viewing them as complementary components of narrative explanation—a form of explanation very common in history.

I. TWO VERSIONS OF CONTINGENCY

Recall Gould's thought experiment:

> You press the rewind button and...go back to any time and place in the past.... [A]ny [such] replay of the tape would lead evolution down a pathway radically different from the road actually taken.

I will call this the "unpredictability" notion of contingency. It figures prominently in *Wonderful Life*. But if that is what contingency is all about, then it is difficult to see why Gould drew so much inspiration from the Frank Capra movie from which he took his title (WL 287–89). In the film, *It's a Wonderful Life*, George Bailey's guardian angel, Clarence, shows the despondent George what an important difference he had made in so many people's lives, by letting him see how sad life in Bedford Falls would have been without him. This is not replaying the tape of life in Bedford Falls

---

Sticklebacks: Size, Shape, and Habitat Use Efficiency," *Ecology*, LXXIV (1993): 699–709; Schluter, "Experimental Evidence that Competition Promotes Divergence in Adaptive Radiation," *Science*, CCLXVI (1994): 798–801; and H.D. Rundle, L. Nagel, J.W. Boughman, and Schluter, "Natural Selection and Parallel Speciation in Sympatric Sticklebacks," *Science*, CCLXXXVII (2000): 306–08.

from the same point in the past, however; this is replaying the tape from a very different point—without George! The difference that George made is not really captured by the unpredictability interpretation of contingency; it is much better captured by another version of contingency that Gould articulates. I will call this the "causal-dependence" version:

> Historical explanations take the form of narrative: E, the phenomenon to be explained, arose because D came before, preceded by C, B, and A. If any of these earlier stages had not occurred, or had transpired in a different way, then E would not exist (or would be present in a substantially altered form, É, requiring a different explanation). Thus, E makes sense and can be explained rigorously as the outcome of A through D....
>
> I am not speaking of randomness..., but of the central principle of all history—*contingency*.... A historical explanation [rests] on an unpredictable sequence of antecedent states, where any major change in any step of the sequence would have altered the final result. This final result is therefore dependent, or contingent, upon everything that came before—the unerasable and determining signature of history (WL 283).[5]

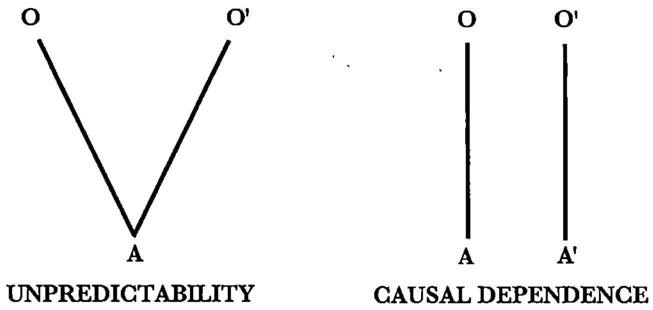
On the face of it, this would seem a very different version of contingency. On the first version, we get different, unpredictable outcomes from the same or indistinguishable prior states (Figure 1, left). Which might sound somewhat mysterious. Whereas on the second version, the particular outcome depends strongly on which particular states preceded it (Figure 1, right). Different prior states lead to different outcomes. Which might sound somewhat trivial.

Another way of putting the difference is that on the unpredictability version, the occurrence of a particular prior state is *insufficient* to bring about a particular outcome, whereas on the causal-dependence version, the occurrence of a particular prior state is *necessary* (or strongly necessary) to bring about a particular outcome. (Yet another way of putting the difference is to note that these two versions correspond to the two colloquial uses of the term "contingent"—"contingent *per se*," and "contingent upon"—the former indicating unpredictability, and the latter causal dependence.)

---

[5] In the passage just cited, Gould states that "E *had* to arise, as a consequence of A through D" (my emphasis). Here he seems to contradict his own position on the unpredictability of evolutionary outcomes. In this paper, I am treating this apparent contradiction as a misstatement.

KP2566



UNPREDICTABILITY   CAUSAL DEPENDENCE

Figure 1: The two versions of historical contingency. (a) The unpredictability version: from the same starting point A (A for ancestral), evolution leads unpredictably to different outcomes O and O′. (b) The causal-dependence version: depending on the starting point, A or A′, evolution leads to O or O′. See text.

These two versions of historical contingency are clearly compatible. We might even think of them as complementary components of a combined interpretation, according to which:

> a historically contingent sequence of events is one in which the prior states are *necessary* or strongly necessary (causal-dependence version), *but insufficient* (unpredictability version) to bring about the outcome.

Such a combined notion of contingency figures prominently in the literature of philosophy of history and historiography. Gould, himself an accomplished historian and philosopher of history, frequently alluded to this by referring to "historical explanation" and "historical narrative" (for example, in the previous quotation). He might profitably have drawn more on this literature in clarifying his position; but he did not. I will discuss his views in connection with this literature briefly in my conclusion.

In addition to the complementary roles that these notions of contingency play in philosophy of history, they also play complementary roles in Gould's critique of the all-sufficiency of selectionist reasoning, to which I will turn shortly. But the best way to make sense of their roles in Gould's critique of pan-selectionism is to continue to consider them separately. For, as I will explain, Gould presents two different critiques of selectionism, corresponding to the two versions of contingency. This way of proceeding is also useful when considering subsequent tests of Gould's views, which are sometimes aimed at one, and sometimes at the other version or component of contingency. So I will back up and consider each version and its bearing on selectionist thinking in greater detail, beginning with the unpredictability notion.

II. THE UNPREDICTABILITY VERSION OF HISTORICAL CONTINGENCY

Gould first raised the possibility of replaying life and getting different outcomes from the same starting point as a way of understanding why, of many diverse multicellular animals that arose in the early Cambrian period (570 mya), only a small number persisted during the apparent mass extinction that followed. Why those few? Why, for instance, did *Pikaia*—the first known chordate and possibly an ancestor of vertebrates, including humans—persist, rather than *Opabinia, Wiwaxia, Hallucigenia, Anomalocaris* or so many of the other wonderful but extinct forms preserved in the Burgess Shale?[6]

One possibility is that the winners (including our ancestors) were much better adapted to their environments. Indeed, they were so much better adapted, that if we could put all the same Cambrian ancestors back into the same Cambrian environments, we would get the same winners over and over.

Another possibility is that the decimation that followed the Cambrian explosion was more like a "lottery" among forms that were equally well, if differently, adapted, and hence had equal chances of survival (WL 50, 239). In this case, if we replayed the tape we might end up with descendants of *Opabinia, Wiwaxia, Hallucigenia,* or *Anomalocaris* instead, and perhaps *we* would not be around to watch the tape play out.[7]

Generalizing, when Gould speaks of replaying the tape, with the unpredictability version of contingency in mind, he imagines what would be the evolutionary outcomes if initially indistinguishable ancestors were placed in indistinguishable environments (presumably, if the environments change, they would change at the same times and in the same ways). To the extent that natural selection within a particular environment is sufficient to determine which forms prevail, then under these circumstances, the same outcomes would prevail over and over. In denying that the same outcomes will result, Gould is not suggesting that the outcomes are inexplicable. Rather, he is denying that selection alone is sufficient to guarantee one particular outcome. Some other factors that distinguish the lineages or groups of lineages must be responsible.

---

[6] See the illustrations throughout *Wonderful Life.* See also http://www.nmnh.si.edu/paleo/shale/index.html.

[7] The analogy of "replaying the tape" is similar to an analogy used earlier by Simon Conway Morris in contemplating the same phenomena:

> ...if the clock was turned back so metazoan diversification was allowed to rerun across the Precambrian-Cambrian boundary, it seems possible that the successful body plans emerging from this initial burst of evolution may have included wiwaxids rather than mollusks.

—"The Middle Cambrian Metazoan *Wiwaxia corrugata* (Matthew) from the Burgess Shale and *Ogygopsis* Shale, British Columbia, Canada," *Philosophical Transactions of the Royal Society,* London B CCCVII (1985): 507–82, p. 572; quoted by Gould, WL 238.

What other factors? Gould did not, to the best of my knowledge, offer a systematic account of the sorts of factors that would lead to different evolutionary outcomes given initially indistinguishable ancestral lineages and indistinguishable environments. But he did emphasize one important class of factors that I would like to discuss briefly. And then I will also briefly discuss another class of factors that one might think he had in mind (especially given the Burgess Shale example), but that he explicitly excluded from the category of "historically contingent" factors.

In his discussions of contingency in the unpredictability sense, Gould often invoked Charles Darwin's distinction between what is predictable about evolutionary outcomes on the basis of general laws, and the unpredictable "details...[which are] left to the working out of what we may call chance."[8] What is predictable, from Darwin's point of view, is that evolution by natural selection will result in well-adapted forms. What is unpredictable is whether evolution by natural selection will result in this particular adaptive form or that. And the reason is what Darwin called "accidental" variation, and what we call "random" variation or random mutation.

To predict what particular forms life will take, we would need to know more than just the generalization that adaptively more valuable variations and combinations of variations will be selected; we would also need to know what particular variations will arise, and in what particular order. Two initially identical populations inhabiting identical environments will not have simultaneous, identical mutational histories, and as a result they may diverge evolutionarily.

For example, imagine two populations homogeneous for a gene X_1. Suppose the environment of each population changes, but in precisely the same respects. Mutations occur in each population, mostly disadvantageous. But ultimately an advantageous mutation—say, X_5—occurs in the first population and increases in frequency. In the second population, a different advantageous mutation—X_{12}—occurs and increases in frequency.

It might be argued that this divergence would be temporary at best. Eventually, X_{12} will occur in the first population and will increase in frequency if it is more advantageous than X_5. Or else X_5 will occur in the second population and increase in frequency if it is more advantageous than X_{12}. In either case, the two populations will eventually converge evolutionarily.

---

[8] Darwin to A. Gray, 22 May 1860, in Darwin, *The Correspondence of Charles Darwin, Volume 8, 1860*, F. Burkhardt, J. Browne, D.M. Porter, and M. Richmond, eds. (New York: Cambridge, 1993), pp. 223–24. See also WL 290; and Gould, "Boyle's Law and Darwin's Details," *Natural History*, CIV (1995): 8–11, 68–71, especially pp. 69–71.

But even if the two populations underwent the same mutations in the long run—at one particular locus—and even if one of those mutations was significantly more advantageous than the rest, the populations might still diverge, depending on differences in their mutational histories at other loci. The problem gets much more interesting when we consider the timing and order of mutations at multiple loci, and the possibility (known as "epistasis") that whether a mutation at one locus is advantageous depends on what genes it is teamed up with at other loci.

To see how this might happen, return to the previous example, and imagine what is happening at a second locus. Suppose the initially identical populations are homogeneous for X_1 and Y_1 (Figure 2).

Following the mutations at locus X, X_5Y_1 prevails in the first population, while $X_{12}Y_1$ prevails in the second population. Now suppose that Y_4 occurs in each population. Suppose further that the X_5Y_4 combination is less advantageous than X_5Y_1, so that X_5Y_1 continues to prevail in the first population; whereas, the $X_{12}Y_4$ combination is more advantageous than $X_{12}Y_1$, so that $X_{12}Y_4$ increases in frequency in the second population.[9]

Add more epistatic loci to the example and the difference in evolutionary outcomes will be even greater.

One of Gould's—and Darwin's—favorite examples of the unpredictability of evolution (based on the unpredictability of mutations and mutational order) concerns the evolution of the ichneumonids

---

[9] Here is an admittedly fictional, but hopefully not overly contrived example of the abstract scenario just described. (This example is inspired by the Losos study to which I will turn shortly.) Suppose there are two very similar species of lizards living on different islands, but in very similar environments, spending most of their time on the ground and running to escape predators. Suppose also that the environment on each island changes in the same way, such that it is advantageous (let us imagine equally advantageous) to occupy either trees or bushes. The bushes, with their narrower stems and branches, are more difficult to run along, but suppose that they have some other advantage that makes them as advantageous a habitat as trees. Because of random mutation and differences in the order of mutation on the two islands, the lizards on one island evolve to become tree dwellers, while on the other island, they evolve to become bush dwellers. The tree lizards might well continue to run to escape their predators; that is, tree dwelling and running for protection might be a good combination; whereas the bush lizards, which run at best haphazardly, might evolve alternative means of protection, for example, cryptic coloration, along with the instinct to freeze whenever a predator is near.

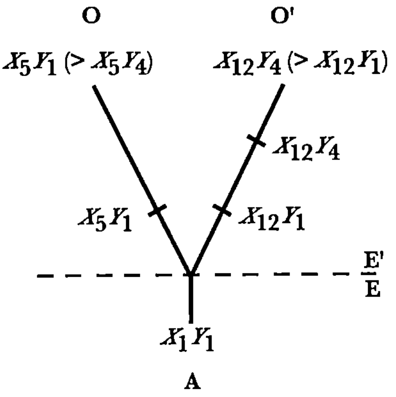


Figure 2: Evolutionary divergence from the same starting point, due to changes in mutational order and epistasis. See text.

(parasitic wasps). One might well predict ahead of time that the reproductive behaviors and life histories of organisms would be adaptive. But could one have predicted the succession of mutually adaptive mutations leading to the ichneumonids, which insert their eggs into the bodies of living caterpillars where the eggs develop into larvae that consume the caterpillar from the inside-out, eventually emerging (à la *Alien*, the movie), and then living as pupae attached to the empty caterpillar carcass, until they emerge as wasps to repeat the hoary cycle?

The ichneumonids were especially telling for Darwin. While he could possibly imagine that God had designed the general laws of nature, like evolution by natural selection, he strongly resisted the thought that God might have foreseen, and have either intended or allowed that particular variations would occur in a particular order leading, for instance, to the repugnant ichneumonids. Darwin preferred to think that God had let variation occur randomly. That is why he spoke of the "details, *whether good or bad*, [which are] left to the workings out of what we may call chance" (my emphasis). For Gould, as for Darwin, those "details" were a source of considerable unpredictability.

The class of factors that one might have expected Gould to invoke in order to account for contingency in the unpredictability sense is a kind of factor that evolutionary biologists have long invoked to this end, namely, sampling error. The paradigmatic case involves heterozygotes—organisms that have two different genes for a trait instead of two copies of the same gene—just by chance passing down more of one gene than the other to their offspring, thus resulting in an evolutionary change—a "random drift"—in the frequencies of those two genes.

Sampling error leading to random drift of gene frequencies is usually depicted with a common ancestral starting point—a common gene frequency—and a common environment, but with disparate evolutionary outcomes. Figure 3 is a typical example.

This looks a lot like the unpredictability version of contingency (just turned on its side). But Gould explicitly excluded sampling error and random drift from the category of historical contingency. In his own words, he distinguished what he called "stochastic effects" like random drift from "history," "the contingencies of history," and "the happenstances of history."[10] This is a somewhat confusing distinction, since random mutation and mutational order, which Gould acknowledged as sources of historical contingency, are themselves stochastic processes. Gould did not provide any sort of analytical grounds for this distinction.

I will have a bit more to say about this later. But I cannot go on without noting that this distinction—between stochastic effects like random drift due to sampling error, on the one hand, and historical contingency, on the other hand—does not square very well with Gould's own account of the persistence of some forms rather than others following the Cambrian explosion. According to Gould, that outcome was highly contingent and could well have turned out otherwise because it had so little to do with differences in adaptive fit and was more like a lottery. Which sounds a lot like sampling error at the level of lineages. And Gould was certainly friendly to the idea of analogous evolutionary processes taking place at multiple levels of the genealogical hierarchy.[11]

Anyway, to summarize concerning the unpredictability version of contingency—before turning to the causal-dependence version—it is not tantamount to any mysterious sort of indeterminism. It does not deny that evolutionary outcomes can be explained. It merely denies that evolution by natural selection is sufficient to guarantee the same evolutionary outcome, even given initially indistinguishable ancestral lineages and indistinguishable environments, and even excluding stochastic processes like random drift.

---

[10] Gould and D.S. Woodruff, "History as a Cause of Area Effects: An Illustration from *Cerion* on Great Inagua, Bahamas," *British Journal of the Linnean Society*, XL (1990): 67–98.

[11] For example, Gould, *The Structure of Evolutionary Theory* (Cambridge: Harvard, 2002).

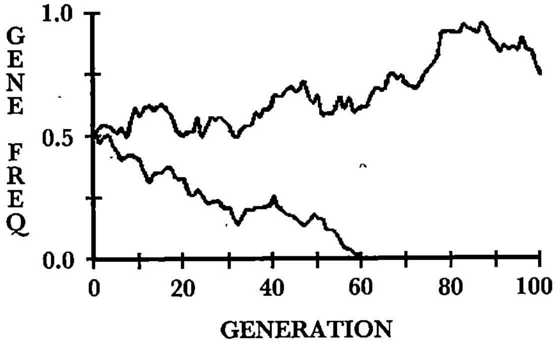


Figure 3: Random drift of gene frequencies leads to two different outcomes from the same starting point (no selection, population size 100; generated by Donald Alstad's *Populus* program). See text.

III. THE CAUSAL-DEPENDENCE VERSION OF HISTORICAL CONTINGENCY

While the first version of contingency emphasizes the unpredictability of evolutionary outcomes, the second version emphasizes the causal dependence of outcomes on prior circumstances, which might seem obvious enough, but which is intended in a way that is not trivial.

Consider that a marble in the bottom of a bowl could have rolled there from anywhere inside the lip (Figure 4).

Its present position is, in this sense, not dependent on its particular prior position and carries no trace of its past. The physics of the situation leads inexorably from any starting point inside the lip to the same position at the bottom.

Evolution, Gould argued, is not like that. Nothing, not even natural selection, leads inexorably from any starting point to the same outcome—not even in indistinguishable environments. Present (and future) forms of lineages are highly dependent upon their prior ancestral states. The main reason for this, according to Gould, involves what he called "historical constraints." These are factors "internal" to the ancestors of a lineage that restrict the range of variations upon which natural selection can act to modify further the lineage. In Gould's terms, historical constraints "restrict the freedom of natural selection to establish and control the direction of evolution" (*ibid.*, p. 1028). Historical constraints ensure that future forms of a lineage are inscribed with the "unerasable signatures" of past forms—that is, unerasable by natural selection (WL 283).

For example, neither whales nor fish have legs, and from an adaptive point of view, it is sufficiently clear why. But the four-legged ancestors of whales provided very different variations for natural

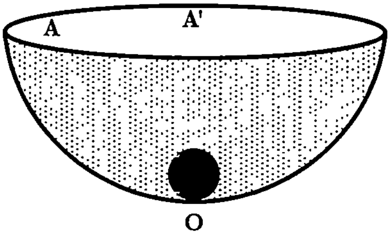


Figure 4: An outcome that carries no trace of its past. See text.

selection to act upon than did the legless ancestors of fish. And this difference in variations available to natural selection is evident in the different outcomes. Whales may no longer have legs, but they carry around rudimentary pelvises that natural selection has not completely erased. The exact outcome of whale evolution is thus historically contingent in the sense of being causally dependent on the ancestral state of whales.[12]

Before moving on to the empirical tests of Gould's views, it is worth pausing to note that, while the unpredictability and causal-dependence versions of historical contingency make different points, they are similar in at least one respect. That is, they make the same general point vis-à-vis evolution by natural selection. Together, they rule out the sort of evolutionary inexorability portrayed in Figure 5. That is, they rule out the idea that evolution by natural selection in a specified environment will result in a predictable outcome *regardless of ancestral starting point.* (And what is more, this noninexorable nature of evolution has nothing to do with sampling error and random drift.)

---

[12] When Gould discussed the causal dependence of evolutionary outcomes on prior states, he usually (as I noted earlier) emphasized historical constraints on the range and direction of variation. This is, at least on the face of it, a very different source of contingency than random mutation and mutational order, which he emphasized in connection with the unpredictability of evolutionary outcomes. But it is worth noting briefly that random mutation and mutational order are not only a source of unpredictability, but also a reason for causal dependence of outcomes on prior states. Consider the scenario in Figure 2 that I discussed earlier. Here the difference in outcomes is dependent on the difference in prior states, X_5Y_1 and $X_{12}Y_1$, a difference that was unpredictable given ancestral state A (X_1Y_1) and the environment inhabited by each population. Moreover, it might be argued that as the initial differences are compounded by further differences in the order of mutation, and epistatic interactions with yet other traits, then the descendants of each lineage will eventually differ enough that they will offer quite different ranges of variation for further evolution by natural selection. In this way, random mutation and mutational order might give rise to different historical constraints.

ALIPORE
CAMPUS

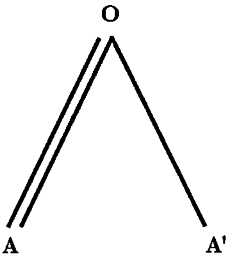


Figure 5: What the two versions of historical contingency, together, rule out: namely, that the same evolutionary outcome would result twice from the same ancestral starting point, and even from different starting points, presumably due to the overwhelming influence of natural selection. See text.

Nonetheless, interesting and important differences between the two versions of historical contingency remain. The sources of contingency in the unpredictability sense—random mutation and mutational order—render the future somewhat independent of the past. While the sources of contingency in the causal-dependence sense—historical constraints on variation—inscribe the future with the past.

Differences between the two versions are also important to keep in mind in considering attempts to test Gould's view(s) on contingency.

IV. TESTING GOULD'S VIEWS: A MACROEVOLUTIONARY "NATURAL EXPERIMENT"

Toward the end of *Wonderful Life*, Gould mused, regretfully, that of course the tape can never be replayed. There can never be a controlled experiment to confirm or disconfirm the contingency of an evolutionary outcome. However, as he more cheerfully noted, nature sometimes presents us with fairly well designed evolutionary experiments featuring closely related—and hence similar, if not indistinguishable—lineages inhabiting fairly similar environments over long periods of time (WL 297–99). Let us now consider one of those "natural experiments," as analyzed by the Losos group (*op. cit.*). And then finally I will consider something like an actual replaying of the tape as designed and analyzed by the Travisano group (*op. cit.*)

Losos and colleagues were primarily concerned to put the causal-dependence version of historical contingency to the test. And this is reflected in the first sentence of their paper: "The theory of historical contingency proposes that unique past events have a large influence on subsequent evolution" (*op. cit.*, p. 2115). But their study actually bears on both versions of contingency, and there are indications in their article that this was not entirely inadvertent. For example, their title, "Contingency and Determinism in Replicated Adaptive Radiations..." suggests that they were addressing the unpredictability interpretation, by pursuing the question whether the outcomes of replicated radiations are contingent, or predictable. The first sentence of their abstract suggests the same:

> The vagaries of history lead to the prediction that repeated instances of evolutionary diversification will lead to disparate outcomes even if starting conditions are similar (*op. cit.*, p. 2115).[13]

Losos and his colleagues studied the evolutionary diversification of lizards of the genus *Anolis* on four islands in the Caribbean: Cuba, Hispaniola, Jamaica, and Puerto Rico. No single species occurs on more than one island, but the same "ecomorphic" *types* of species occur over and over (an "ecomorph" is a type of species that is adapted to a particular niche). For example, each island has its own "twig ecomorph" species of *Anolis* that is well adapted to living on and navigating narrow twigs. These lizards have relatively short legs and well developed toe pads (they also have a slow, searching form of foraging behavior, and rely primarily on cryptic coloration combined with stillness for defense against their own predators). And each island has its own "trunk-ground ecomorph" species that is well adapted to life on tree trunks and the ground. These lizards have longer limbs and less well developed toe pads (and they run and grab moving prey, and run to escape their own predators). There are also two other ecomorphs on each island, adapted to life on two other parts of trees. The adaptive fit of the various ecomorphs to their preferred habitats has been well studied, for example, by "chasing" different types of lizards along rods of different diameters, and measuring their speed and agility.[14]

---

[13] The first two sentences of the body of the Losos paper suggest that the two versions are logically related—that the unpredictability interpretation follows from the causal-dependence version:

> The theory of historical contingency proposes that unique past events have a large influence on subsequent evolution. A corollary is that repeated occurrences of an evolutionary event would result in radically different outcomes (p. 2115).

If the corollary is that replays from identical starting points in identical environments lead to different outcomes (leave the "radically" aside), then of course it does not follow. However, Losos (personal communication, August 3, 2004) confirms that the second sentence should be interpreted as making a point more in keeping with the first—something like: replays from different starting conditions, even in *similar* environmental conditions, will lead to different outcomes. His point is that, in the real world, environments are at best similar, never identical.

[14] Losos and B. Sinervo, "The Effects of Morphology and Perch Diameter on Sprint Performance of *Anolis* Lizards," *Journal of Experimental Biology* CXLV (1989): 25–30. For

How did it come to be that each island has its own representative of each of these four ecomorphs? One extreme possibility is that each island was originally colonized by the same ecomorph. And then from that type of ancestor, the other three ecomorphs evolved independently on each island. This sounds somewhat like the opposite of Gould's replay expectation, in that basically the same evolutionary outcome would have occurred four times (that is to say, on four islands) from basically the same ancestral starting point, and in basically the same environmental conditions. In this case, one could even say that the *same four outcomes* (the same four ecomorphs) would have occurred *four times* (on four islands), under basically the same environmental conditions.

Another extreme scenario is the Gouldian possibility that each ecomorph arose only once, as one of many, equally satisfactory ways of adapting to a Caribbean environment. Descendants of each original ecomorphic species then made their way around to the various Caribbean islands (by dispersal or vicariance) and survived there nicely.

The Losos group attempted to distinguish between these extreme possibilities by reconstructing and analyzing the evolutionary relationships of the species in question (based on mitochondrial DNA sequences). On the first hypothesis, the species living on any one island, although representing four different ecomorphs, should be more closely related to each other than to species on any other island (Figure 6, top). On the second hypothesis, the species within any one ecomorphic group, while inhabiting separate islands, should be more closely related to each other than to species within any other ecomorphic group, even species on the same island (Figure 6, bottom).

In fact, their phylogenetic analysis revealed that species on each island were more closely related to each other than to species on any other island, strongly suggesting that each ecomorphic species arose on each island: four replays of the evolution of the same four ecomorphs.

With one interesting difference: the evolutionary relationships of the ecomorphs were different on each island (Figure 7). Losos and his colleagues suggested that these differences could be understood as differences in the temporal order of evolution on each island,

---

an overview of these ecomorphs, see E.E. Williams, "Ecomorphs, Faunas, Island Size, and Diverse End Points in Island Radiations," in R.B. Huey, E.R. Pianka, and T.W. Schoener, eds., *Lizard Ecology: Studies of a Model Organism* (Cambridge: Harvard, 1983), pp. 326–70. For photos of the ecomorphs, see P.H. Harvey and L. Partridge, "Different Routes to Similar Ends," *Nature*, CCCXCII (1998): 552–53; and Todd Jackman's Anole page at http://www87.homepage.villanova.edu/todd.jackman/anolis/anolis.html.

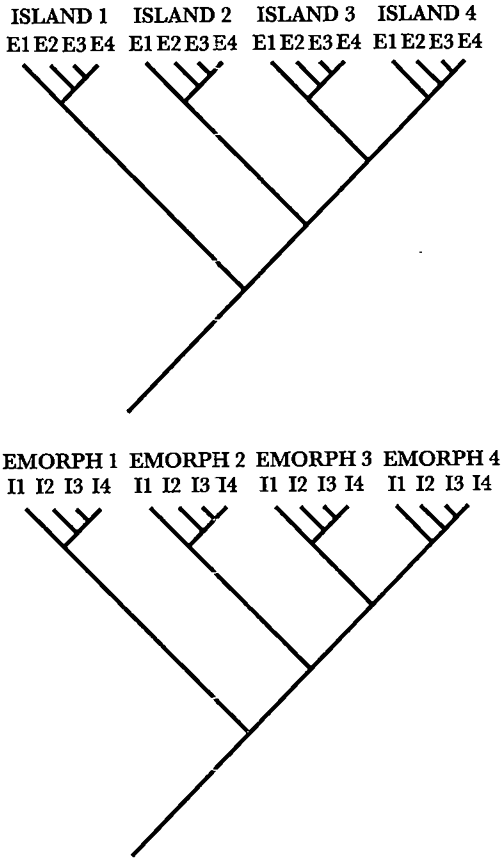


Figure 6: Hypothetical phylogenetic scenarios for investigating the role of historical contingency in the Losos study (much simplified by the author). (a) From the same ecomorphic ancestor, the same four ecomorphs (E1–E4) evolved independently on each island. (b) Each ecomorph arose once; descendants of each ecomorph then eventually colonized each island (I1–I4). See text.

which in turn could be due to contingency in the causal-dependence sense. As they reasoned, the niches (or ecomorphic opportunities) available at any one time on an island depend on which ecomorphs have already arrived or evolved there.[15]

---

[15] Although Losos stresses that the inference from phylogeny to temporal order is not straightforward (personal communication, August 3, 2004).

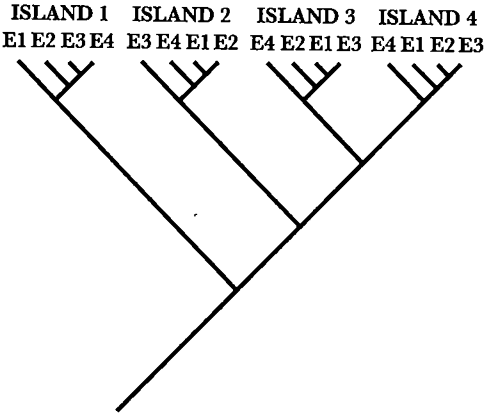


Figure 7: Phylogenetic results of the Losos study (much simplified by the author). See text.

So the Losos study bears on the causal-dependence version of contingency, since they were addressing the extent to which differences in outcome are attributable to different ancestral states. But their work also bears strongly on the unpredictability interpretation of contingency, since their phylogenetic analysis is consistent with the possibility that, from basically the same starting point (the same ancestral ecomorph), the same outcome (the same derived ecomorphs) repeatedly result.

It is worth mentioning one other respect in which the Losos study addressed and ruled out historical contingency in the causal-dependence sense. That is, they raised the possibility that the ancestral lineages that originally colonized the islands were either negatively constrained from producing any ecomorphs other than the set of ecomorphic variants seen there, or positively constrained to produce just those types. In this case, the repeated evolutionary outcomes would have been contingent—in the causal-dependence sense—on the prior state of the ancestral lineage and could not be explained solely in terms of adaptive fit. They ruled out this possibility on the grounds that the colonists presumably arrived from the mainland, and on the mainland there is a wider range of ecomorphs than is found on the islands.

The Losos study was not discussed in print by Gould. But Losos recollects hearing second hand, through a *Science* reporter, that Gould was not moved by their results:

> My recollection is that his response was that the phenomena he was talking about pertained to events deep in evolutionary time (perhaps

referring explicitly to the Cambrian Explosion) and that our lizard work was on recently diverged events.[16]

V. TESTING GOULD'S VIEWS: A MICROEVOLUTIONARY, LABORATORY EXPERIMENT

Given Gould's reaction to the Losos group's *macroevolutionary* study, it is perhaps not too surprising that he had no detailed response to the Travisano group's *microevolutionary* experiment. But, for reasons that I will discuss, Gould might also have felt that he and they were talking past each other somewhat.

While the Losos group focused on the causal-dependence version of contingency, Travisano's group explicitly focused on the unpredictability interpretation. They designed and ran experiments to mimic as closely as possible actual replayings of the evolutionary—albeit microevolutionary—tape. As the Travisano group explained their intent,

> S. J. Gould has argued for the great importance of historical contingency. He has presented a gedanken experiment of "replaying life's tape" to test the repeatability of evolution and thereby evaluate the roles of adaptation, chance and history. Of course, one cannot perform an actual experiment on the grand scale envisioned by Gould, but one can perform rigorous experiments, of shorter duration and in simpler environments, to quantify the roles of adaptation, chance, and history in evolution. Instead of replaying life's tape sequentially, one can achieve the same objective by doing an experiment in which replicate populations are propagated simultaneously (*op. cit.*, p. 87).

This sounds promising. Recall that Gould also distinguished between adaptation, "stochastic" factors and historical contingency. So this experiment might seem to be right on the mark. However, the Travisano study did not distinguish between "chance" and "history" in the same manner that Gould distinguished between "stochastic" factors and "historical contingency." Among the sources of historical contingency, Gould included mutation and mutational order, and he contrasted historical contingency with stochastic effects like random drift due to sampling error. Travisano's group, in a more traditional manner, included in the category "chance" *both* sampling error leading to random drift, *and* random mutation and mutational

---

[16] Personal communication, October 15, 2002. Perhaps Gould had the then-recent Losos study in mind when he proposed a qualified version of the replay metaphor in 1990: "What would happen if we could rewind the tape of life to some moment in the fairly distant past and simply let it run again?"—Gould, *The Individual in Darwin's World: The Second Edinburgh Medal Address* (Edinburgh: University Press, 1990), p. 7.

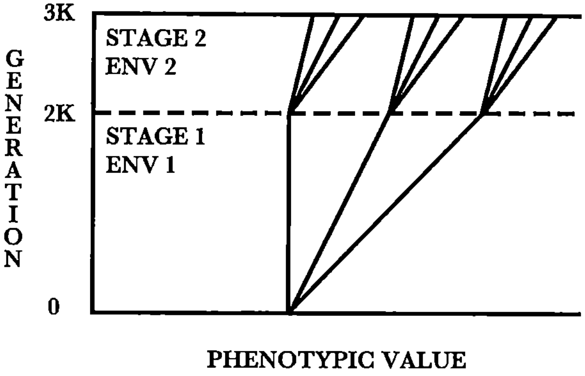


Figure 8: Two stages of the experiments reported in the Travisano study. See text.

order. (Although, as we shall see, in further studies Travisano, Lenski, and their colleagues tried to tease apart the contributions of sampling error and random mutation in ways that Gould would probably have appreciated.) Moreover, they invoked "history" only in those cases where different evolutionary outcomes resulted from different ancestral states (*op. cit.*, p. 87), which corresponds to Gould's causal-dependence version of historical contingency. It was the unpredictability version of historical contingency, however, that they had set out to investigate.

I will discuss two of the four experiments that the Travisano study reported. Each of these experiments had two stages (Figure 8).

In the first stage, they cloned 12 copies of a single genotype of *E. coli* to establish 12 genetically identical populations. Then they grew these populations separately in a glucose-limited environment for 2,000 generations, and measured the derived phenotypes that resulted (I will discuss the phenotypes in question shortly). Following that stage, they cloned 3 copies of one individual from each of the 12 populations to produce 36 populations (of 12 different genotypes), and grew these separately for 1,000 generations in a maltose-limited environment, and then measured the derived phenotypes.

They interpreted the results with the help of the following analytical scenarios, beginning with possible outcomes of the first stage (Figure 9; *op. cit.*).[17] Note that they did not address the role of history

---

[17] I have redrawn these scenarios somewhat in order to make them more comparable with the other figures in this paper. I do not believe that I have altered their content.

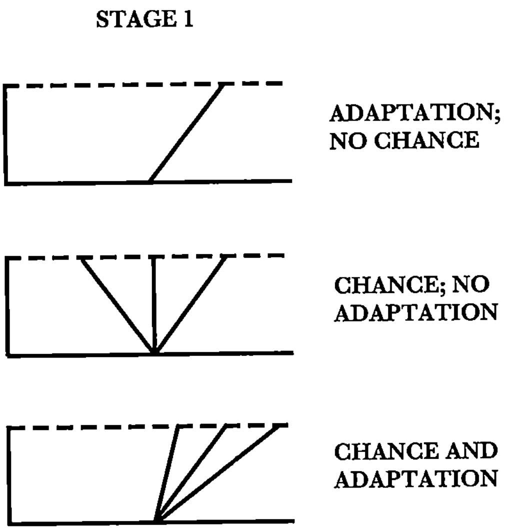


Figure 9: Analytical scenarios for investigating the roles of "adaptation," "history," and "chance" in the first stage of the experiments reported in the Travisano study (redrawn slightly in order to make them more consistent with other representations in the paper). NB: Travisano's group does not address the role of history in the first stage. (a) From the same ancestral state, in the same environment, the same form is derived each time; this is a sign of adaptation alone with no role for chance. (b) From the same ancestral state, in the same environment, a range of forms around the ancestral mean is derived; this is a sign of chance, with no role for adaptation. (c) From the same ancestral state, in the same environment, a range of forms around a new mean is derived; this is a sign of adaptation plus chance. See text.

in the first stage. Indeed, given what they meant by "history," there was no history in the first stage to address.

In the first scenario (Figure 9, top), from the same ancestral state, in the same environment, the same particular form is derived every time. This, they argued, would be best interpreted solely in terms of adaptation—selection in favor of a greater value of the phenotypic variable—with no role for chance (*op. cit.*).

In the second scenario (Figure 9, middle), from the same ancestral state, in the same environment, a range of forms around the mean of the ancestral form are derived. This purely dispersing effect, they argued, would be best interpreted solely in terms of chance, with no role for natural selection (*op. cit.*). Recall that they included in the

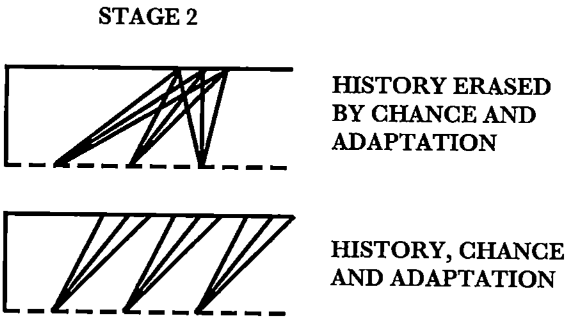

STAGE 2
HISTORY ERASED BY CHANCE AND ADAPTATION
HISTORY, CHANCE AND ADAPTATION

Figure 10: Analytical scenarios for investigating the roles of "adaptation," "history," and "chance" in the second stage of the experiments reported in the Travisano study (redrawn slightly in order to make them more consistent with other representations in this paper): (a) "an initial effect due to history is eliminated by subsequent effects due to chance and adaptation," (b) "an initial effect due to history is maintained with subsequent effects due to chance and adaptation superimposed." See text.

category "chance" not only random drift due to sampling error, but also random mutation, which Gould associated with historical contingency in the unpredictability sense. So "chance" for them overlaps one of Gould's sources of historical contingency.

The third scenario (Figure 9, bottom), they argued, is best interpreted in terms of selection in favor of an optimal phenotype, together with the dispersing effects of chance (*op. cit.*).

Moving now to the analytical scenarios for interpreting the second stage of the experiment (Figure 10).

In the first scenario of the second stage (Figure 10, top), from widely different starting points, a much narrower range of outcomes results. The Travisano group suggested that in this case, "an initial effect due to history is eliminated by subsequent effects due to chance and adaptation" (*op. cit.*). Recall that they associated "history" with differences in ancestral states; history has a role when differences in evolutionary outcomes are due in part to differences in ancestral states. Clearly history as such has no role in this scenario.

In the second scenario (Figure 10, bottom), the Travisano group suggested that, "an initial effect due to history is maintained with subsequent effects due to chance and adaptation superimposed" (*op. cit.*).

Again, the sort of "history" that the Travisano group are looking for with this sort of experiment, and this sort of analysis, is historical contingency in the causal-dependence sense. It was the unpredictability version of historical contingency, however, that they reported to be the focus of their study.

The actual results of their experiments are extremely interesting. They studied two phenotypes, cell size, which they had prior reason to believe was only weakly correlated with fitness, and fitness itself as measured by ability to compete for resources.

I will focus on the results of the fitness experiment. At the end of the first stage, in the glucose-limited environment, the 12 populations had very similar fitness values suggesting (not surprisingly) a significant role for adaptation. However, when those 12 derived populations were *initially* placed in the maltose-limited environment (at the beginning of stage 2), their fitnesses were found to diverge considerably, although these differences were subsequently erased, suggesting again a significant role for adaptation.

But what about those initial differences in fitness in the second, maltose-limited environment? This finding was investigated further in a separate paper, which concluded that considerable genetic differences underlay the similarities in fitness that had developed in the glucose-limited environment, and that the different genotypes represented equally adaptive outcomes within that environment.[18] Brief mention was made of the extent to which random mutation and mutational order might be involved in the differences that occurred at the genetic level.[19]

This finding quite clearly bears on Gould's unpredictability notion of historical contingency, since at least at the genetic level, different outcomes resulted from identical ancestral states in identical environments. Other studies by the Lenski group delved more deeply still into the role of random mutation and mutational order in the unpredictability of evolutionary outcomes.[20] Moreover, in these studies, random mutation and mutational order were sometimes referred to not only as chance events, but also as historically contingent factors—as for example when Lenski and Travisano referred to "the crucial role of chance events (historical accidents) in adaptive evolution" (*ibid.*, p. 6813).

---

[18] Travisano, F. Vasi, and Lenski, "Long-Term Experimental Evolution in *Escherichia coli.* III: Variation among the Replicate Populations in Correlated Responses to Novel Environments," *Evolution,* XLIX (1995): 189–200.

[19] See their reference to G.S. Mani and B.C. Clarke, "Mutational Order: A Major Stochastic Process in Evolution," *Proceedings of the Royal Society of London B,* CCXL (1990): 29–37.

[20] Lenski and Travisano, "Dynamics of Adaptation and Diversification: A 10,000 Generation Experiment with Bacterial Populations," *Proceedings of the National Academy of Sciences of the United States of America,* XCI (1994): 6808–14; and P.A. Johnson, Lenski, and F.C. Hoppensteadt, "Theoretical Analysis of Divergence in Mean Fitness between Initially Identical Populations," *Proceedings of the Royal Society of London B,* CCLIX (1995): 125–30.

Although Gould never commented in detail on the Travisano study, he did refer to it briefly, in a footnote, in his last book, *The Structure of Evolutionary Theory* (*op. cit.*). There he also discussed (at greater length) the study by Lenski and Travisano that I just mentioned and will now describe in more detail. The Lenski and Travisano experiment corresponded to the first stage of the Travisano experiment, extended from 2,000 to 10,000 generations in the glucose-limited environment. Somewhat surprisingly, the fitness values of the 12 populations, which were similar after 2,000 generations, were still *only similar* after 10,000 generations. Why after 10,000 generations in exactly the same environment had not all of the populations evolved the same ability to compete for the resources available in that environment? Lenski and Travisano suggested that the residual differences in fitness reflected underlying genetic differences, which had in turn resulted from the different order of mutations occurring in the twelve populations, compounded by epistasis. They concluded that, "our experiment demonstrates the crucial role of chance events (*historical accidents*) in adaptive evolution" (*op. cit.* p. 6813, my emphasis).

Again, this would seem to bear quite directly on Gould's unpredictability version of contingency. Interestingly, Lenski and Travisano did not cite Gould's views on historical contingency in this particular paper. But Gould certainly saw the connection himself. As he reported their results:

> The general path of increase [for example, in fitness] followed the same trajectory in all populations, but with fascinating differences of both form and genetics in each case [presumably referring to the genetic differences underlying the similar fitnesses]—a remarkable commentary, at such a small and well-controlled scale, of the roles of detailed contingency and broad predictability in evolution.[21]

The "remarkable commentary" was not further articulated.

Apparently, sometime just before the 2002 book went to print, Gould also added a footnote concerning the Travisano experiments, which he took as support for the significance of contingency, but without elaborating why.[22] He did not refer to the ways in which the Travisano group distinguished between adaptation, chance and history. This would not have been something that he could have discussed briefly![23]

---

[21] Gould, *The Structure of Evolutionary Theory*, p. 932.

It is interesting that Gould's discussion of Lenski and Travisano, and his footnote on the Travisano experiments, was in the context of their bearing on punctuated equilibrium—not contingency *per se.* He seemed most interested in the patterns of punctuated evolution in their results. Perhaps this helps to explain why Gould seemed to have a fairly positive view of the Lenski group's microevolutionary studies, relative to his slightly negative view of the Losos group's macroevolutionary studies. Recall his reported critique of the Losos group: that their study did not extend sufficiently deep into evolutionary time.

VI. CONCLUSION

By way of conclusion, I would like to return briefly to two issues, and raise a third anew. First, what are we to make of the two versions of historical contingency? Do we need both? Or does one version have a better claim to the title? There is no question that in ordinary language, "contingent" is used to refer to unpredictable outcomes ("contingent" *per se*) and also to the causal dependence of particular outcomes on prior events (as in "contingent upon"; see, for example, *Merriam-Webster*). But perhaps more importantly, there is a tradition in historiography and philosophy of history of viewing these two notions of contingency as complementary components of historical narratives. W.B. Gallie often made this point:

> In general, the possibility of prediction figures in stories very much as it does in most of our thinking about everyday life. There is a dominant sense of alternative possibilities: events in train are felt to admit of different possible outcomes—particularly those events that count, that evoke praise or blame, that deserve to be recorded, that could be the pivot of a good story. But side by side with this there is the recognition that many events, or aspects of events, are predictable either exactly or

---

[22] Concerning the seeming *non sequitur* that concludes Gould's footnote, I think it makes more sense if you replace "Apparently" with "Moreover" (*The Structure of Evolutionary Theory*, pp. 932–33).

[23] Travisano believes that an historical accident of sorts may explain why Gould discovered the Travisano experiment at such a late stage in the writing of his book. Due to a publisher's snafu, the Travisano article was not actually listed in the table of contents of the issue of *Science* in which it appeared. Another article was listed in its place. So Gould may have overlooked it, and then may not have been able to find it even when it was first brought to his attention (personal communication, October 25, 2003; this confused me as well, the first time I tried to track down the paper).

> approximately. But, although recognized, this predictable aspect of life is, so to speak, recessive or in shadow. It is in contrast to the generally recognized realm of predictable uniformities that the unpredictable developments of a story stand out, as worth making a story of, and as worth following.
>
> We should notice here that perhaps of greater importance for stories than the predictability relation between events is the converse relation which enables us to see, not indeed that some earlier event necessitated a later one, but that a later event required, as its necessary condition, some earlier one. More simply, almost every incident in a story requires, as a necessary condition of its intelligibility, its acceptability, some indication of the kind of event or context which occasioned or evoked it, or, at the very least, made it possible. This relation, rather than the predictability of certain events given the occurrence of others, is the main bond of logical continuity in any story.[24]

In a nutshell: if the outcome of a historical narrative were predictable from the outset, or at any point, there would be no need to include any further events—that is, no need for the rest of the narrative. So the early and intermediate events in a narrative should be insufficient to make sense of the outcome. Of the intermediate events that *are* included, at least some should have a causal bearing on the outcome—otherwise, why include them? So a good, self-contained narrative should be constructed out of events that are insufficient but necessary for the outcome.

But even if there is good reason to stick with a dual notion of contingency, rather than viewing one or the other version as *the* correct account, still, it helps to state clearly which version is at issue in discussions and empirical investigations of contingency. This helps, for instance, in appreciating how the Losos study bears on more than just the causal-dependence version of historical contingency, and how the Travisano experiments bear on more than just the unpredictability version.

The second point to which I would like to return briefly concerns the distinction between historical contingency and chance. How should that distinction be drawn? Some will find the Travisano distinction more compelling than Gould's. Recall that Travisano's study included random variation and random drift in the "chance" category, while Gould put random variation in the "historical contingency" category, leaving random drift in the "stochastic" category. The emphasis on "random" may suggest a natural link that Gould artificially severed. It may be that he was thinking along the following lines (I am not trying to defend his distinction here; for the time being I just want to understand it). Random variation and historical constraints, for all their differences, both have to do with the *origin and direction* of variation, whether random or constrained. Evolution by natural selection and sampling error leading to random drift, for all their differences, both have to do with the *differential persistence of existing variation*. By emphasizing the importance of historical contingency relative to evolution by natural selection and random drift, then, Gould may have intended to emphasize the need to take into account the origin and direction and not just the persistence of variation in explaining evolutionary outcomes.

---

[24] Gallie, *Philosophy and the Historical Understanding* (New York: Schocken, 1964), p. 26.

But who really cares whether, for instance, random mutation and mutational order go into the category of "historical contingency," or into the category of "chance"? Why do we not just investigate, for instance, the relative extents to which random mutation and mutational order, versus random drift, versus adaptive evolution are responsible for evolutionary outcomes? Interestingly, while Lenski and Travisano (*op. cit.*) include random mutation, mutational order, and random drift all in the "chance" category, nonetheless they distinguish between the effects of random mutation and mutational order on the one hand, and random drift on the other, and argue that the former is more important in explaining the divergence of populations in the 10,000 generation experiment.

Why place so much emphasis on the distinction between historical contingency and chance? I suspect we will be in a better position to answer that question when we understand why, beginning in the mid-late nineteen seventies and early eighties, there was so much concern to emphasize the importance of "history," not only in evolutionary biology, but also in other areas of biology like ecology, biogeography, conservation biology and systematics, and in the social sciences as well (see footnote 1).

Finally, it is important to consider what the issues concerning Gould's views of historical contingency have in common with so many other issues in biology, and what will make the issues concerning contingency difficult to resolve long after we have sorted out the conceptual issues at stake. This is the problem of extrapolating from the importance of an evolutionary factor in one or a few, or even many cases, to its overall importance. What does one study, or one hundred, or even one thousand tell us about the overall importance of historical contingency in either or both of the senses that Gould articulated? Of course, Gould did write that "*any* replay of the tape would lead evolution down a pathway radically different from the one taken" (WL 51; my emphasis). In principle, one negative case would undermine his position (at least with regard to one of the versions of contingency). But we should not forget that Gould was also the co-author of that famous line in "Spandrels of San Marco:"

> In natural history, all possible things happen sometimes; you generally do not support your favored phenomenon by declaring rivals impossible in theory. Rather, you acknowledge the rival, but circumscribe its domain of action so narrowly that it cannot have any importance in the affairs of nature. Then, you often congratulate yourself for being such an ecumenical chap.[25]

Surely, like so many disputes in biology, the issue will become one of the *relative significance*—rather than the *universality*—of historical contingency.[26] Losos charitably acknowledges that so far the evidence bearing on Gould's position is "anecdotal."[27] What sorts of studies, short of a complete tally of evolutionary episodes, will give us more than anecdotal insight into the overall importance of historical contingency?[28]

JOHN BEATTY

University of British Columbia

---

[25] Gould and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London B*, CCV (1979): 581–98, p. 585.

[26] Beatty, "Why Do Biologists Argue Like They Do?" *Philosophy of Science S*, LXVI (1997): 231–42.

[27] Personal communication, September 27, 2002.

[28] And even if historical contingency is somehow determined to play a role in most every case, there will still be questions about how big a role relative to simultaneously coacting factors. See S. Mitchell, "On Pluralism and Competition in Evolutionary Explanations," *American Zoologist*, XXXII (1992): 135–44.

COMMENTS AND CRITICISM

DEFENDING FRANKFURT'S ARGUMENT IN DETERMINISTIC CONTEXTS: A REPLY TO PALMER

In earlier work, we have taken issue with a nearly unanimous assumption regarding Harry Frankfurt's argument that moral responsibility does not require alternative possibilities.[1] The assumption is that the examples figuring in Frankfurt's argument, prior-sign examples (to be illustrated below), cannot assume a deterministic relation between states of an agent prior to her free actions and the subsequent actions themselves. Why? Appeal to such examples, it is held, begs the question against the incompatibilist. But we maintain that the matter is not so clear. We agree that in certain dialectical contexts the assumption of a deterministic relation *is* illicit, and so in these contexts, Frankfurt's argument succeeds only if the assumption is avoided. However, we argue that there are other dialectical contexts in which it is legitimate to use these examples to help support the conclusion that moral responsibility does not require alternative possibilities. In the latter dialectical contexts, we do not mean to claim that such examples demonstrably establish Frankfurt's conclusion. We only mean to make the modest claim that the examples aid in making a credible case for the relevant conclusion, one that perhaps is open to dispute. Our point is simply that, in the dialectical contexts at issue, the charge of begging the question is unwarranted.

Recently, David Palmer has rejoined that our line of reasoning is not cogent; he submits that, contrary to what we claim, in the pertinent dialectical contexts, Frankfurt cases that assume determinism *are* question-begging.[2] In this paper, we focus attention on Palmer's central objection to our position, which we find intelligent and probing.[3] We believe that a demonstration as to where Palmer has

---

[1] Ishtiyaque Haji and Michael McKenna, "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," this JOURNAL, CI, 6 (June 2004): 299–314.

[2] Palmer, "New Distinctions, Same Troubles: A Reply to Haji and McKenna," this JOURNAL, CII, 9 (September 2005): 474–82.

[3] Although we shall not pursue these matters here, we wish to note that Palmer misrepresents a number of crucial points that we develop. Consider just one instance where Palmer distorts our position. He reports us as claiming that in a certain dialectical context, *any* incompatibilist who demands an indeterministic setting for Frankfurt examples makes the demand in bad faith (477). But this was not what we

0022-362X/06/0307/363–72

gone wrong will help to cast yet further light on the delicacy of the dialectical contexts at issue.

I. FRANKFURT'S ARGUMENT AND PRIOR-SIGN CASES

Here is a typical example of a prior-sign Frankfurt case, a crucial feature of which is a sign that "detects" a freely willed action temporally prior to its occurrence:

> Mork believes that in his circumstances it is wrong for him to lie to prevent embarrassment. In spite of this, he decides to lie and lies anyway. As he has no excuse or justification for lying, it seems he is deserving of blame for lying. Unbeknownst to Mork, though, he could not have refrained from lying due to the presence of Mindy who has the power to read and control Mork's mind. Mindy wields this power partly in virtue of possessing the following knowledge. Had Mork been about to refrain from lying, he would have displayed some involuntary sign—a neurological pattern, N^*, in his brain—whereas if he had been about to lie, he would have displayed a different neurological pattern, N. Had Mindy detected N^*, she would have interceded in Mork's deliberations via direct stimulation of Mork's brain and would, in this way, have caused Mork to lie. But Mindy detects N, the reliable sign for Mindy that she need not show her hand at all. As it happens, Mork lies on his own.[4]

---

claimed. We proposed, rather, the following. *Some* incompatibilists hold that the sole basis for their position is grounded in the thought that determinism rules out alternative possibilities. *Other* incompatibilists might agree with this basis for incompatibilism, but argue that other reasons also license an incompatibilist conclusion, one being that in a determined world our actions derive from sources over which we have no control. Now in one dialectical context—the broad dialectical context—the Frankfurt Defender makes a concession to the incompatibilist by agreeing to a key incompatibilist demand. This incompatibilist insists that, if she is to be persuaded by a Frankfurt case, the case be one in which a libertarian free agent is morally responsible for her conduct despite not having alternatives. But if libertarian free and responsible agency on a certain incompatibilist construal is grounded *only* in alternative possibilities, then this incompatibilist demand is tantamount to asking Frankfurt to prove the possibility of a contradiction: "Provide me with a case in which an agent with alternative possibilities has no alternative possibilities." Hence, we argued, the only incompatibilists who are entitled to demand that Frankfurt work with a case of libertarian free agency are those incompatibilists who have some basis for their incompatibilism that is not anchored in determinism's ruling out alternatives. These incompatibilists can make the relevant demand without doing so in bad faith, and we believe that this is an important insight in understanding the subtleties of the debate. Since the force of Palmer's thoughtful, core objection to our position stands independently of issues such as these, we set them aside.

[4] "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," p. 300.

Frankfurt's argument against alternative possibilities as a condition of moral responsibility builds upon a case such as this one. Because Mork lies on his own, in the absence of Mindy's intervention, it seems highly reasonable that Mork acts freely and is morally blameworthy for lying despite not having alternative possibilities.

II. THE DILEMMA DEFENSE

One of the most potent objections against Frankfurt's argument takes the form of a dilemma.[5] Either the relationship between the prior sign Mork exhibits and his subsequent decision is deterministic or it is indeterministic. In the former case, the examples are able to guarantee a single course of action, but it is in dispute that the agent is free and responsible. In the latter case, while the agent is free and responsible, it is false that the agent cannot do otherwise.

To explain: suppose, on the first horn, that the relationship is deterministic. Then, so the objection goes, the cases, right at the outset, beg the question against incompatibilists who believe that determinism is incompatible with free action and moral responsibility. With the assumption of a causally deterministic relation, it would be true that Mork has no pertinent alternatives. Once Mindy detected *N*, she could rest assured that Mork would certainly lie. And were she to detect *N**, she would have insured the same. So in this case, the example really does guarantee that Mork lies. But incompatibilists would insist that, given the assumption of a deterministic relation, it is false that Mork acts freely and is morally responsible for his decision to lie.

Suppose, on the second horn, that the relationship is indeterministic. Then though it may well be true that Mork is morally responsible for his decision to lie, it would be false that Mork could not have decided to do otherwise even with the counterfactual intervener—Mindy—in place. Should Mindy detect *N*, given the assumption of indeterminism, it is consistent with *N* that Mork lie or that he not lie.

Refer to those who believe that this dilemma undermines Frankfurt's argument *Dilemma Defenders*; and call those who think that Frankfurt cases do undercut the requirement of alternative possibilities for moral responsibility *Frankfurt Defenders*.

---

[5] See, for example, Carl Ginet, "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing," *Philosophical Perspectives*, x (1996): 403–17; Robert Kane, *Free Will and Values* (Albany: SUNY, 1985); and David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *Philosophical Review*, CIV (1995): 247–61.

III. TWO DIALECTICAL CONTEXTS: BROAD AND NARROW

Most Frankfurt Defenders have simply assumed that the only way to respond to Dilemma Defenders is to take on the second horn of the dilemma, the indeterministic horn. Indeed, each of us has independently resisted Dilemma Defenders by resisting the second horn.[6] However, we also agree with John Martin Fischer that Frankfurt Defenders should not be so quick to concede the first, deterministic horn to the Dilemma Defender.[7] We believe that attention to the first horn is fruitful, since it draws into relief important assumptions about the dialectic that otherwise go unnoticed. We turn to a treatment of the first horn.

To assess fully this dilemma we have proposed that, among other things, it is important to distinguish two interpretations of the first horn (or two dialectical contexts regarding the first horn). On the broad interpretation of the first horn, Frankfurt Defenders mean to persuade incompatibilists, already theoretically *committed* to a position on the compatibility of determinism and moral responsibility, that if determinism does rule out the freedom moral responsibility requires, it is *not* by virtue of ruling out alternative possibilities. The committed incompatibilist believes that a determined agent is not free or morally responsible in the first place. So if Frankfurt cases are to convince such an incompatibilist, on pain of not begging the question, the Frankfurt Defender must offer a Frankfurt case that does *not* assume determinism. On the broad interpretation of the first horn, what is objectionable about the assumption of determinism—the reason that it is question begging—is due to the very assumption of determinism itself. *It* is in opposition to the incompatibilist thesis. Palmer seems to agree with us on this score. He has qualms, though, concerning the narrow interpretation.

On the narrow interpretation of the first horn, both Frankfurt Defenders and Dilemma Defenders aim to persuade those who are *undecided* about the free will problem, those who have not yet committed to compatibilism or incompatibilism. In this context, Frankfurt Defenders mean to convince the "undecided voter" that if determinism is incompatible with the freedom moral responsibility requires, it is not because determinism rules out alternative possibilities. By contrast, the Dilemma Defenders mean to convince the same group of undecided voters that if determinism is incompatible with the freedom moral responsibility requires, a credible basis for that incompatibility is that determinism rules out alternative possibilities, and nothing in the Frankfurt cases should convince us otherwise. Given this dialectical context, on the deterministic horn, the Dilemma Defender is not entitled to the complaint that the prior sign in a Frankfurt case assumes a deterministic relation *simpliciter*. The undecided voter should not be expected to be swayed by this assumption. Thus, in this dialectical context, it is moot whether the assumption of determinism is objectionable. On the other hand, relying on the narrow interpretation, the Dilemma Defender might think it reasonable to object that, if determinism *already* rules out alternative possibilities (as incompatibilists claim that it does), then there is no work for the counterfactual intervener to do. Dilemma Defenders might demand that to insure that an important question is not begged, Frankfurt Defenders must produce a Frankfurt case in which determinism plays no role at all in ruling out alternatives. Thus, the question begging feature of the deterministic horn in this dialectical context, it is alleged, is due not to the mere fact of determinism, but rather to the fact that alternatives get ruled out for reasons other than those introduced by Frankfurt's counterfactual machinery.

---

[6] Haji, *Moral Appraisability* (New York: Oxford, 1998), pp. 16–41; "The Principle of Alternate Possibilities and a Defeated Dilemma," *Philosophical Explorations*, IX, 2 (2006): 179–202; and McKenna, "Robustness, Control, and the Demand for Morally Significant Alternatives," in Widerker and McKenna, eds., *Moral Responsibility and Alternative Possibilities* (Aldershot, UK: Ashgate, 2003), pp. 201–18.

[7] Fischer, "Recent Work on Moral Responsibility," *Ethics*, CX (1999): 93–139.

We claimed, though, that in this dialectical context, the charge of begging the question is too strong. The uncommitted voter can appreciate what work the counterfactual intervener is doing in a Frankfurt case that assumes determinism without begging any important questions. Of course, the key point is this: Is there a legitimate way for the Frankfurt Defender to illustrate the irrelevance of alternative possibilities in a case in which, on certain views, alternative possibilities are already ruled out for controversial reasons (namely, determinism)? We held that there is a way to illustrate this point, and we reasoned roughly as follows:

> There are prima facie plausible prior-sign Frankfurt cases in which the counterfactual intervener, on its own, rules out alternatives. The intervener's ruling out alternatives is irrelevant to any judgment concerning the agent's acting freely or being morally responsible. Thus, if this way of expunging alternatives makes no difference to judgments of free action or responsibility, then this provides some reason for thinking that the ruling out of alternatives by determinism, barring special considerations to believe otherwise, should not make a difference either.[8]

---

[8] This strategy was first defended by Fischer in "Recent Work on Moral Responsibility." Our earlier piece was offered as a limited defense of Fischer's contribution.

Palmer summarizes our response on behalf of Frankfurt Defenders to Dilemma Defenders in this way:

(1) Determinism and the counterfactual intervener are individually sufficient for ruling out alternative possibilities.
(2) The counterfactual intervener way of ruling out alternative possibilities is irrelevant to judging whether an agent is free and has moral responsibility.
(3) So, from (1) and (2), the determinism way of ruling out alternative possibilities is irrelevant to judging whether an agent is free and has moral responsibility (479–80).

This formulation of our response overstates our case. Given (1) and (2), we merely held that a plausibility thesis could be developed to *suggest* that alternative possibilities are irrelevant to judging whether an agent is free and morally responsible. Regardless, Palmer's main objection applies with equal force even to the weaker thesis that we did defend. So consider his compelling challenge.

IV. PALMER'S OBJECTION AND A REPLY

Palmer rejects line (1) above. He writes:

> The counterfactual intervener can only rule out alternative possibilities in conjunction with an assumption that determinism holds. Put another way, the counterfactual intervener *qua* counterfactual intervener is not sufficient to rule out alternative possibilities.... To see that this is so...return to...[the] presentation of the dilemma.... For the counterfactual intervener (Mindy, in the example) to pick up on the "prior sign" as a reliable indicator of the agent's (Mork's) choice and action, determinism must hold, since indeterminism would deny a reliable relationship between the two. Imagine Mork decides not to lie. In order for Mindy to intervene successfully, without Mork's knowledge, determinism must hold, since Mindy detects the prior sign and activates different neural pathways sufficient for Mork actually to lie. However, if we assume an undetermined environment, there will be no determined relationship between the prior sign and the choice. So, if Mindy is to intervene, she will have to wait until Mork has already, in some sense, chosen to lie before being able to alter Mork's neural mechanisms. The Frankfurt Defender wants, of course to rule out such a possibility for this is an instance of a robust alternative possibility. So, for the counterfactual intervener to rule out alternative possibilities, which would involve him picking up on the prior sign, the relevant Frankfurt case must assume determinism holds (480).

The crux of Palmer's probing objection distils to the following. In a prior-sign Frankfurt case, the presence of the counterfactual intervener (Mindy), *only* in conjunction with an independently sufficient condition for Mork's choosing to act as Mindy desires (the condition that Mork's choice is causally determined), ensures that Mork has no alternatives. What Palmer's objection requires, really, is the stronger contention that it is not possible for there to be a case in which the counterfactual intervener eliminates alternatives without the assumption of determinism. For if there were such a case, the claim that the counterfactual intervener *qua* counterfactual intervener is sufficient to rule out alternative possibilities would be vindicated. And then surely we could simply wed one of these cases to a deterministic context.

A first indication that all is not well with Palmer's concern is brought to light by recognition that Frankfurt cases can be used to persuade *Traditional Compatibilists* that free action and responsibility do not require alternatives. Such compatibilists endorse a principle that requires alternative possibilities for moral responsibility, but they interpret "can" and "could have done otherwise" in a compatibilist fashion. On one such interpretation, for instance, to say that an agent, *s*, can do action, *a*, is to say that *s* would do *a* if *s* wanted to do *a* or if *s* tried to do *a*. In suitably crafted Frankfurt cases, the counterfactual machinery, a component of which is the counterfactual intervener, suffices to eliminate what we may call "compatibilist alternatives."[9] Thus, if eliminating (compatibilist) alternatives in this manner makes no difference to ascriptions of free action or moral responsibility, this gives us *some* reason to believe that there can be prior-sign Frankfurt cases in which the counterfactual machinery on its own, independently of determinism, rules out alternatives. When we say that there can be such cases in which the counterfactual machinery *on its own* rules out alternatives, we should be understood as saying that the counterfactual machinery, together with other background conditions that do not include the assumption of determinism, rules out alternatives.

To construct such a case, we begin with a Frankfurt example that Derk Pereboom develops in which the agent who corresponds to Mork is a *libertarian free agent*.[10] In Pereboom's example, the prior sign marks the presence of a condition that is necessary but not sufficient for any alternative to the desired course of action. In this necessary-condition prior-sign case, the agent freely makes some decision, such as the decision to lie or to cheat. But she could not arrive at an alternative decision. This is because, if she were to satisfy some merely necessary (and not sufficient) condition to make an alternative decision, an intervener would guarantee that she decide as she actually does. Here are the details of Pereboom's case:

---

[9] See, for example, Haji, *Deontic Morality and Control* (New York: Cambridge, 2002), p. 68; and McKenna, "Does Strong Compatibilism Survive Frankfurt-Style Counter-Examples?" *Philosophical Studies*, XCI, 3 (1998): 259–64.

[10] See Pereboom, "Alternative Possibilities and Causal Histories," *Philosophical Perspectives*, XIV (2000): 119–37, *Living without Free Will* (New York: Cambridge, 2001), "Source Incompatibilism and Alternative Possibilities," in Widerker and McKenna, eds., pp. 185–99, and "Defending Hard Incompatibilism," *Midwest Studies in Philosophy*, XXIX (2005): 228–47.

> Tax Evasion (2): Joe is considering whether to claim a tax deduction for the substantial local registration fee that he paid when he bought a house. He knows that claiming the deduction is illegal, that he probably won't be caught, and that if he is, he can convincingly plead ignorance. Suppose he has a very powerful but not always overriding desire to advance his self-interest regardless of the cost to others, and no matter whether advancing his self-interest involves illegal activity. Crucially, his psychology is such that the only way that in this situation he could fail to choose to evade taxes is for moral reasons. (The phrase *failing to choose to evade taxes* is meant to encompass not choosing to evade taxes and choosing not to evade taxes.) His psychology is not, for example, such that he could fail to choose to evade taxes for no reason or simply on a whim. In addition, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to these moral reasons. He can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, Joe could, with his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). More generally, Joe is a libertarian free agent. But to ensure that he choose to evade taxes, a neuroscientist now implants a device, which, were it to sense the requisite level of attentiveness, would electronically stimulate his brain so that he would choose to evade taxes. In actual fact, he does not attain this level of attentiveness, and he chooses to evade taxes while the device remains idle.[11]

In this scenario, it is highly plausible that the counterfactual machinery is sufficient to eliminate robust alternatives (of course, the machinery expunges alternatives without assuming determinism). Now, minimally modify the case so that Joe is a *determined* agent. It will still be true that Joe can fail to choose to evade taxes only if he attains a certain level of attentiveness to apt moral reasons, and that he can secure this level of attentiveness voluntarily, although "voluntarily" is now to be construed in some compatibilist fashion. On one such construal, for example, it must be that to attain voluntarily a level of attentiveness, Joe be appropriately responsive to suitable reasons. We see no motivation at all to deny that, with the minimal modification of "subtracting" libertarianism from the scenario and "adding" determinism instead, the counterfactual machinery suffices to rule out alternatives.

---

[11] "Source Incompatibilism and Alternative Possibilities," p. 193, notes omitted.

To tie some ends together, in a prior-sign Frankfurt case, two factors (determinism and the counterfactual machinery) each eliminates alternatives. The machinery is irrelevant, on its own, to moral responsibility. Were it true that it is not possible for the machinery itself, without the added assumption of determinism, to expunge alternatives, then we would not be able to draw the moral that we wish: we would not be able, with propriety, to infer that since the counterfactual intervener way of ruling out alternative possibilities is irrelevant to judging whether an agent is morally responsible, this suggests that the determinism way of ruling out alternatives might also be irrelevant. But we have claimed that it is possible for the machinery itself, minus the assumption of determinism, to exclude alternatives. We admit that to marshal support for this claim, we appealed, first, to a Frankfurt case in which the prior sign is a necessary condition on the agent's availing himself of any robust alternatives. We do not, though, see why this is in any way illicit, if the appeal is meant to debunk the view that it is not possible for the counterfactual machinery on its own (absent any assumption of determinism) to expunge alternatives.

We conclude that Palmer has failed to discredit our view that, in relevant dialectical contexts, prior-sign Frankfurt cases that assume determinism do not beg the question against the Dilemma Defender. We began this paper by promising to cast further light on the dialectical delicacies at issue in the debate over Frankfurt's argument. Notice that once the dialectical context is specified as the narrow one, and once it is thereby acknowledged that the dispute is to be settled by convincing the fair and theoretically uncommitted voter, it is irrelevant whether the example that is meant to turn the trick is deterministic or indeterministic. It is only in what we have identified as the broad dialectical context that the issue of determinism matters. And as we observed in our previous paper, in that case, the background assumption is that the Frankfurt Defender should engage the incompatibilist by beginning the debate on the incompatibilist's turf, granting her, so to speak, the assumption of incompatibilism.[12] But for those who think that the relevance of alternative possibilities should be settled either prior to or independently of the truth of incompatibilism, it is unclear why the Frankfurt Defender should begin the debate on foreign soil. Why not instead settle the matter on neutral terrain? Seen as such, in the narrow dialectical context, Frankfurt's argument has a sort of simplicity and innocence to it that can otherwise be overlooked: Regardless of the truth or falsity of determinism, freedom over alternatives seems not to be what is at issue when assessing the freedom pertaining to an agent's moral responsibility. What matters turns simply on *what an agent has done.*

---

[12] "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," p. 314.

ISHTIYAQUE HAJI
University of Calgary

MICHAEL MCKENNA
Florida State University

THE JOURNAL OF
PHILOSOPHY

BACK ISSUE SALE

Complete volumes
are now available
at the following discounted prices
in order to reduce our inventory:

Volumes	1 through 98	$10 each
Volumes	99 through 101	$25 each

For more information, contact Pamela Ward,
Business Manager, at pcw1@columbia.edu

Visa, Mastercard, American Express and
PayPal are accepted through our website at
www.journalofphilosophy.org

Kinesis: A Graduate Journal in Philosophy

Kinesis believes that the philosophical work of graduate students constitutes a substanti contribution to the debate on contemporary philosophical and socio-political issue Kinesis was founded to offer a forum to graduate students in all fields to express the philosophical views in a critical yet open-ended context. We believe that when creati thought renders explicit the relation between philosophical concepts and contempora problems, basic philosophical ideas are alive and working.

Contributions in any area of investigation will be considered, provided th establish the viability of arguments and conclusions of the author. Response articl concerning works previously published in Kinesis are encouraged. All submissio should include a cover letter and a copy of the submission saved in Microsoft Word. A inquiries concerning the status of essays under review must be submitted in writing a directed to the Editor.

– Subscription Rates –

One Volume (two issues)	
Students:	$15.00
General Subscribers:	$20.00
Single Issue:	$10.00

All inquiries should be directed to:
Kinesis
Department of Philosophy
Southern Illinois University at Carbondale
Carbondale, IL 62901
(618) 453-7447
kinesis@siu.edu
http://www.siu.edu/~philos/kinesis/

Ethics

an international journal of social political and legal philosophy

John Deigh, Editor

Founded in 1890, ***Ethics*** is an international journal of moral, political, and legal philosophy presenting scholarly work from a variety of intellectual perspectives, including social and political theory, law, and economics. ***Ethics*** also publishes book reviews and review essays.

Ethics Online

www.journals.uchicago.edu/ET

- Links to author references, tables, and figures
- Easy, printable PDF format
- Full-text articles searchable by keyword

Quarterly. ISSN: 0014-1704

e-TOC

Sign up for e-TOC, the FREE tables of contents alerting service at ***www.journals.uchicago.edu/ET***

To Subscribe:
Online: *www.journals.uchicago.edu/ET*
E-mail: *subscriptions@press.uchicago.edu*
Phone: toll-free (877) 705-1878 or (773) 753-3347
Fax: toll-free (877) 705-1879 or (773) 753-0811

3/05

The University of Chicago Press
Journals Division • Box 37005 • Chicago, IL 60637

REVUE
INTERNATIONALE DE
PHILOSOPHIE

Direction: Michel Meyer

c/o Université Libre de Bruxelles, C.P. 188
50, avenue Franklin Roosevelt / B-1050 Bruxelles (Belgique)
Tél. 32 2 650 25 61 / Fax 32 2 650 45 61
e-mail : tsalis@ulb.ac.be

2-2006 / volume 61 / n° 236

L'ontologie

DENIS MIÉVILLE

Logique, Ontologie et Ontologie

ROBERTO POLI

Levels of reality and the Psychological Stratum

FRÉDÉRIC NEF

Ontologie de l'objet, théorie des propriétés et théorie des ensembles : quelques problèmes et perspectives

MARIO VALENTINO BRAMÈ

String theory : in search of (meta)physical objects

MICHEL WEBER

La virtualité en procès. Relativisation de l'acte et de la puissance chez A.N. Whitehead

PRIX DE CE NUMÉRO : 27,83 EUROS

Prochains fascicules / Forthcoming:

Rawls – Philosophie de la musique
Mondialisation – Aesthetics

CAMBRIDGE

BEST IN SCHOLARSHIP

Psychology and the Natural Law of Reparation
C. Fred Alford
$70.00: Hb: 0-521-86332-5: 182pp

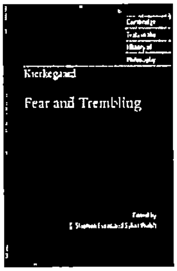
Evolutionary Ethics and Contemporary Biology
Giovanni Boniolo and
Gabriele De Anna
Cambridge Studies in Philosophy and Biology
$75.00. Hb: 0-521-85629-9: 220pp

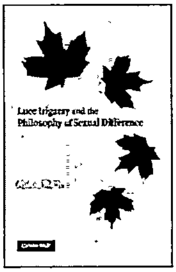
Descartes and the Passionate Mind
Deborah J. Brown
$85.00: Hb: 0-521-85728-7: 244pp

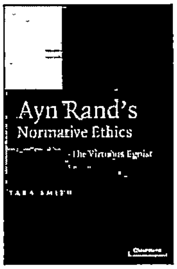
Reasoning in Biological Discoveries
Essays on Mechanisms, Interfield Relations, and Anomaly Resolution
Lindley Darden
Cambridge Studies in Philosophy and Biology
$75.00: Hb: 0-521-85887-9: 360pp

Luce Irigaray and the Philosophy of Sexual Difference
Alison Stone
$80.00: Hb: 0-521-86270-1: 260pp

Goodness and Justice
A Consequentialist Moral Theory
Joseph Mendola
Cambridge Studies in Philosophy
80.00: Hb: 0-521-85953-0: 336pp







Spinoza's 'Ethics'
An Introduction
Steven Nadler
$75:00: Hb: 0-521-83620-4: 300pp
$28.99: Pb: 0-521-54479-3

Kierkegaard: Fear and Trembling
Edited by C. Stephen Evans and Sylvia Walsh
Cambridge Texts in the History of Philosophy
$55.00: Hb: 0-521-84810-5: 190pp
$16.99: Pb: 0-521-61269-1

The Cambridge Companion to Heidegger
Charles Guignon
Cambridge Companions to Philosophy
$75.00: Hb: 0-521-82136-3: 456pp
$29.99: Pb: 0-521-52888-7

Aesthetics and Cognition in Kant's Critical Philosophy
Rebecca Kukla
$75.00: Hb: 0-521-86201-9: 324pp

Rational Choice and Democratic Deliberation
A Theory of Discourse Failure
Fernando Tesón and
Guido Pincione
$80.00: Hb: 0-521-86269-8: 270pp

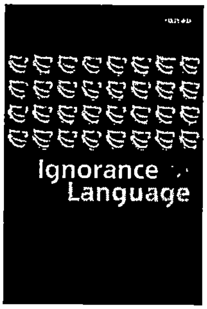
Ayn Rand's Normative Ethics
The Virtuous Egoist
Tara Smith
$80.00: Hb: 0-521-86050-4: 328pp

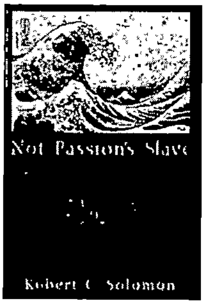
Prices subject to change

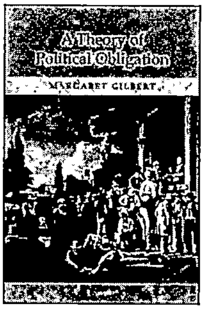
www.cambridge.org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Ignorance of Language
MICHAEL DEVITT

What is linguistics about? What role should linguistic intuitions play in constructing grammars? What is innate about language? Is there a "language faculty?" These questions are crucial to our developing understanding of ourselves; Michael Devitt offers refreshingly original answers.

May 2006 $55 00

New in Paperback!
Not Passion's Slave
Emotions and Choice
ROBERT C. SOLOMON

"The twelve essays by Robert C. Solomon that comprise *Not Passion's Slave* serve as a kind of intellectual memoir of their author, who has, for the last thirty years, been at the heart of a revival of philosophical interest in the emotions."—*Times Literary Supplement*

paper June 2006 $19 95

A Theory of Political Obligation
Membership, Commitment, and the Bonds of Society
MARGARET GILBERT

Margaret Gilbert offers an incisive new approach to a classic problem of political philosophy: when and why should I do what the law tells me to do? Gilbert's accessible book offers a provocative and compelling case in favor of citizens' obligations to the state, while examining how these can be squared with self-interest and other competing considerations.

May 2006 $55 00

New in Paperback!
Reason and Value
Themes from the Moral Philosophy of Joseph Raz
Edited by R. JAY WALLACE, PHILIP PETTIT, SAMUEL SCHEFFLER, and MICHAEL SMITH

The subtlety and power of Raz's reflections on ethical topics—including especially his explorations of the connections between practical reason and the theory of value—make his writings a fertile source for anyone working in this area.

paper May 2006 $34.95

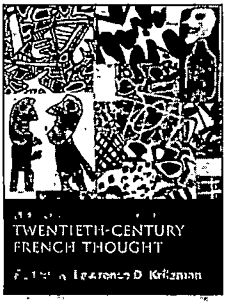
Oxford Studies in Metaphysics
Volume 2
Edited by DEAN ZIMMERMAN

OSM offers a broad view of the subject, featuring not only the traditionally central topics such as existence, identity, modality, time, and causation, but also the rich clusters of metaphysical questions in neighboring fields, such as philosophy of mind and philosophy of science.

May 2006 paper $35.00 cloth $99.00

C O L U M B I A, *Refer to us*

Read book excerpts at www.columbia.edu/cu/cup

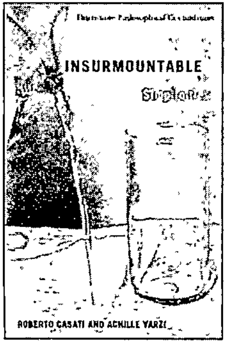


The Columbia History of Twentieth-Century French Thought

Edited by Lawrence D. Kritzman

An indispensable contribution to the intellectual history of France and the twentieth century.

cloth - $75.00



Insurmountable Simplicities

Thirty-nine Philosophical Conundrums

Roberto Casati & Achille Varzi

Elegant, intriguing, and scintillating stories creatively explore a variety of philosophical questions.

cloth - $24.95

Nietzsche and Philosophy

Revised Edition

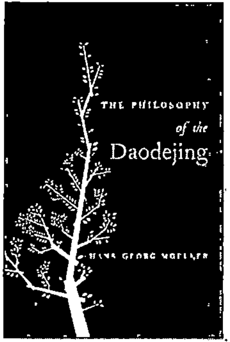
Gilles Deleuze

Foreword by Michael Hardt

European Perspectives: A Series in Social Thought and Cultural Criticism

An influential and controversial study of Nietzsche.

paper - $24.50

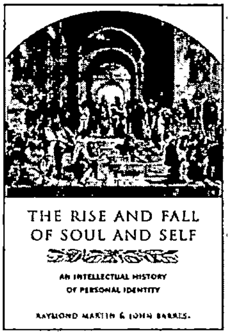


The Philosophy of the Daodejing

Hans-Georg Moeller

An illuminating exploration of a canonical but cryptic work of Daoism and ancient Chinese thought.

paper - $19.50



The Rise and Fall of Soul and Self

An Intellectual History of Personal Identity

Raymond Martin & John Barresi

From Plato and Aristotle to Freud and Foucault, this new history traces the efforts of philosophers and psychologists to understand the soul and the self.

cloth- $29.50

COMING SOON:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy will be adding
a print and online subscription option for institutions
to begin Fall 2006

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

We expect to launch this new feature with our September 2006 issue.
Please consult our website at www.journalofphilosophy.org
for more details.

Calcutta University
14·2·07
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CIII, NUMBER 8
AUGUST 2006

page
373 *Needs (Someone Else's), Projects (My Own), and Reasons* Sarah Buss
403 *Error through Misidentification: Some Varieties* Annalisa Coliva
426 NEW BOOKS: ANTHOLOGIES

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2006

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.

POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2006 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CIII, NO. 8, AUGUST 2006

NEEDS (SOMEONE ELSE'S), PROJECTS (MY OWN), AND REASONS*

I might have chosen a career that permitted me to do more to help people in need. I might have chosen to make more money so that I could spend more on helpful projects; I might have chosen to work at a less time-consuming job so that I could devote more energy to such projects, or to local, national, and international policies that would have more beneficial effects; I might have chosen to make a profession out of helping others. As far as I can tell, these choices are still open to me. I believe I could make any one of them without sacrificing my happiness or settling for a diminished life.[1]

Under these circumstances, I cannot help wondering whether I can really justify the way I have chosen to live my life. I do not want to change course. Indeed, I have no intention of doing so any time soon. But I am less confident than I would like to be that reason is on my side.

Are my reasons for persisting in my chosen career really good enough? Are they good enough, *even if*, as I believe, I could, without great

---

*This paper has benefited from the feedback of many people. I wish to thank Tal Brewer, Toni Koch, Maggie Little, Elijah Milgram, and Connie Rosati for their comments on various written drafts. I am also grateful for having had the chance to discuss the paper with audiences at Richmond University, Kansas State University, the University of Iowa, the University of Michigan, and the University of North Carolina.

[1] I am thus setting to one side the vexed question of whether, and if so, under what circumstances, we can be expected to make such sacrifices. I will not, that is, address the apparent tension between the conviction that "whatever morality requires of us, it should not make our lives unlivable, or too severe" and "the conviction that we must negotiate a decent response to the irrefutable facts of need—of hunger, disease, and poverty" (Barbara Herman, "The Scope of Moral Requirement," *Philosophy and Public Affairs*, xxx, 3 (Summer 2001): 227–56, see p. 228). By defining my problem in such a way that this tension is not at its heart, I avoid being distracted by issues that are irrelevant to my situation, as I understand it, and irrelevant to the situation of most people who are privileged enough to pursue minimally helpful careers. See also notes 7, 33, and 49.

0022-362X/06/0308/373–402

personal sacrifice, spend my time and energy in more helpful ways? I am under no illusion that I was "put on this earth for the sole purpose of serving it or its subjects."[2] I am well aware that it is not possible for me to "make all ends of all others my ends also."[3] My question is: Given that I have sufficient reason not to become anyone's servant, do I have sufficient reason not to pursue a more helpful career?

My uncertainty on this point does not reflect a commitment to any particular moral theory. Rather, independent of any such commitment, I believe that the fact that others need my help is a very strong reason in favor of my helping them, and that this reason is especially strong when their need is especially great—when, for example, their very lives are in jeopardy.[4] This belief poses a problem for me because it naturally prompts the question of whether my reason to help is stronger than whatever reasons I may have to refrain from helping.

Both consequentialist and Kantian moral theories appear to advise me to alter my present course.[5] To be sure, much ink has been spilled in efforts to show that the appearances are misleading. Given the unpredictable entanglements of cause and effect, I might possibly be of more help by maintaining my commitment to philosophy and satisfying the preferences that underlie this commitment. So, too, it is possible that my capacities, or my "effectiveness," as a rational agent would be significantly compromised were I to quit my job.[6] Tempted as I am, however, to believe that these are more than the barest possibilities, I seriously doubt it. Though I know I might be mistaken, it nonetheless seems reasonable to assume that there are many alternative careers that have a more beneficial impact than my own (or at least offer *me* the opportunity to have a more beneficial impact); and it seems reasonable to assume that I could pursue any number of these careers without ceasing to be an effective (indeed, a flourishing[7]) agent. More to the point, I do take both of these things to be true; and it is the implication of these assumptions that concerns me.

---

[2] Susan Wolf, "Morality and Partiality," *Philosophical Perspectives 6: Ethics* (1992): 243–59, see p. 245.

[3] Herman, "The Scope of Moral Requirement," p. 234. As Herman points out, this is impossible, not only because our resources are limited and because people's ends conflict, but because "for everyone to strive to make all ends of all others their ends also would not leave enough of ends that are truly a person's own to form a conception of the happiness to promote" (p. 234).

[4] Though I believe that I have the same reason to help those who cannot read, or whose teeth are decaying, or who are too ill to be "productive members of society," or.... I will focus on the more desperate cases, in part because this permits me to avoid the difficult question of what constitutes a "need." Even if needs that are conditional on various wants are problematic, basic needs—for things without which we cannot have anything else we want—are uncontroversially genuine, if any needs are.

[5] Of course, the Kantian account does not require that we do everything we possibly can to help others. But as far as I can tell, it does not provide a compelling defense of minimally helpful careers. To the contrary, it implies that most of us ought to make far greater efforts to help than we do. For a thoughtful exploration of Kant's account of our imperfect duty to help others, see Thomas Hill, "Beneficence and Self-love: A Kantian Perspective," *Social Philosophy and Policy*, x, 1 (1993): 1–23, and "Meeting Needs and Doing Favors," in *Human Welfare and Moral Worth* (New York: Oxford, 2002), pp. 201–43. See also Herman, "Morality and Everyday Life," Presidential Address, *American Philosophical Association Pacific Division Proceedings*, LXXIV (2000): 29–45, and "The Scope of Moral Requirement."

[6] Herman, "The Scope of Moral Requirement," p. 241. If, as Herman suggests, we are justified in enjoying life at the expense of the welfare of others only insofar as this is

Many who share the assumptions think they raise a problem, not for me, but for the moral theories themselves: any theory that implies that I ought to forego the career of a philosopher has got to be mistaken! I am not so sanguine, however. For, again, it seems to me that my doubts about whether my decision is justified presuppose neither that rational agency is an "end in itself" nor that I ought to do what I can to ensure that the history of the universe is filled with as many valuable things and events as possible. Though it may well be a matter of "common sense" that I have sufficient reason to forego a more helpful career, the arguments for this assumption are as difficult to find outside moral theory as within.

I could try to defend my choice by challenging the assumption that one ought to help others unless one has good reason not to. I could, that is, deny that there is at least *a* (good) reason to help. But I do not want to deny this. To the contrary, like most other philosophers and nonphilosophers, I am convinced it is true. An alternative strategy would be to try to show that the presumption in favor of helping is

---

necessary to our being "successful" (p. 241), "able" (p. 242) agents, who can "appreciate and so correctly evaluate the range of human concerns" (p. 242), then it is hard to see how I can appeal to Kantian morality to justify pursuing a minimally helpful career. Even if "the pursuit of happiness is the organizing principle of our kind of agency" (p. 244), satisfying basic "needs on which agency depends" (p. 243) surely does not require me to be a philosopher—nor, indeed, does the satisfaction of such basic needs require anyone else to pursue a philosophical career. Regarding the appeal to happiness, see also note 7.

[7] Several very interesting studies appear to establish that human beings greatly exaggerate the effect that various welcome or unwelcome changes will have on their happiness. (For a summary of some of these studies, see Jon Gertner, "The Futile Pursuit of Happiness," *The New York Times Magazine* (September 7, 2003): 44, 46–47, 86, and 90.) When people actually experience significant positive and negative changes, they tend, by their own reports, to return very quickly to their earlier level of happiness. When we add to this natural tendency the not unreasonable assumption that people bear some responsibility for the attitudes they take toward their situation, I cannot convince myself that I would cease to be happy if I were to change my career.

relatively easy to override. This strategy is favored by philosophers across a wide range of normative and metaethical views. According to some, our personal interests and concerns have a special normative status which grounds a "prerogative" to give less "weight" to the interests and concerns of others, and so to give less weight to whatever reason(s) we have to help them.[8] Others argue that there are many projects whose value is of a distinctively different kind from the kind at stake in moral judgments, and that there is thus no reason to think that values of the sort realized by a philosophical life are inferior to the value of helping human beings in need.[9]

These arguments sound promising. For one thing, both premises seem true: my commitments do seem to have a special normative significance for me, and there do appear to be many valuable activities whose value is entirely independent of their usefulness. More importantly, the conclusion is of the right sort: it does not depend on establishing that a career in philosophy is really more *useful* than I take it to be. Since my dedication to philosophy is not based on its utility, an adequate defense of this dedication must explain why I am justified in *not* treating my potential to help as the consideration that breaks the tie among the many satisfactory lives I might live.

Attractive though these arguments are, however, in the pages that follow I will review my reasons for concluding that they are not ultimately successful. Whether taken together or singly, the premises to which they appeal do not tip the balance in favor of my decision to continue living a far less helpful life than I might. To be sure, given the extent to which my commitment to philosophy is central to the point of view from which I now evaluate my options, and given that philosophy is a valuable activity, my defense of my career choice is not limited to the rather pitiful observation that "this is what I really want to do and, after all, I can do it without neglecting my family responsibilities or intending to harm anyone." Nonetheless, I will argue, the additional considerations do not add up to an adequate defense. In order to make a case for my choice, I must look elsewhere. I must turn my attention to the heterogeneous collection of rather mundane considerations whose strength it is nearly impossible to assess, especially given the distorting effects of self-interest. These considerations may justify the way I have chosen to structure my life.

---

[8] See Samuel Scheffler,' *The Rejection of Consequentialism* (New York: Oxford, 1982).

[9] See Susan Wolf, "Moral Saints," this JOURNAL, LXXIX, 8 (August 1982): 419–39; and Harry Frankfurt, "The Importance of What We Care About," in *The Importance of What We Care About* (New York: Cambridge, 1988), pp. 80–94, and "Reply to Wolf," in Sarah Buss and Lee Overton, eds., *Contours of Agency* (Cambridge: MIT, 2002), pp. 245–52.

Then, again, they may not. One thing is certain: the best case in support of my choice is bound to be far more fragile than I would like it to be.

This is, I will argue, not merely a verdict concerning the merits of my career choice—or of any such choice. It is inseparable from a more general conclusion about reasons for action. Philosophers who press the two claims mentioned above suggest, in effect, that our reasons not to do more to help others in need are really quite deep. They are as deep as the nature of rational agency and the nature of values. Most of us, these philosophers reassuringly argue, have sufficient reason not to make much effort to help because the "personal" point of view from which every rational agent necessarily deliberates has a special normative significance. Our neglect is further justified, moreover, because helping is just one of many valuable things we can do, because there is no scale on which we can weigh its value against the value of all sorts of other possible endeavors, and because the world would be a far drearier place if it did not contain a wide range of valuable things.

Unfortunately, from what I can tell, these claims yield more comfort than understanding. In particular, they greatly exaggerate the significance of deep facts about reasons and values. This exaggeration has been easy to overlook because the challenge to minimally helpful careers—and to other "projects" that are of relatively little benefit to most human beings, and even less benefit to the many human beings in desperate need—has typically been cast (even by those who press this challenge) as presupposing either the "impersonal," "objective" point of view,[10] or (where it is assumed that there are many incommensurable objective evaluative judgments) the so-called "moral" point of view.[11]

---

[10] A prime example of what I have in mind is Shelley Kagan's characterization of his disagreement with Scheffler. In *The Limits of Morality*, he summarizes his target as follows: "Person's have a point of view. It is a significant fact about us that each of us views the world from a particular perspective. We act out of this perspective, evaluate changes in the world from it, and are affected by those changes at it. It is from this point of view that we undertake plans and projects, and develop interests and desires.... From this point of view persons assess the world, weighing the relative benefits and disadvantages of various actual or potential changes. The evaluation is relative to the particular person, for it is relative to his particular personal perspective; it is a judgment of what is good from his point of view. Typically, it will differ from the agent-neutral evaluation of the impersonal perspective, i.e., an evaluation of what is good objectively. For when persons assess events from their point of view, they assign weights to those events in which they take interest out of proportion to the weights which an impartial perspective would assign." "[T]here are special *subjective* reasons for a given agent to pursue his interests.... [T]hese reasons are directly grounded in the subjective standpoint and thereby differ from reasons based solely on considerations of objective value" (Kagan, *The Limits of Morality* (New York: Oxford, 1989), p. 258 and p. 334; see also pp. 259–70, 335–37). As will soon become obvious, what I object to in both Scheffler and Kagan is the suggestion that if evaluative judgments are made from a given point of view, then they are not judgments about what is good "objectively."

By definition, the "impersonal" or "objective" point of view is external to the "personal" point of view of any individual agent. Accordingly, it can seem that the task of defending minimally helpful careers is as simple as the task of pointing out that we need a *reason* to occupy (or incorporate) this more "objective" point of view (into) ourselves. Why assume that we ought to value what we *would* value from some *external* point of view? Why assume that this point of view is privileged? Why assume that the reasons accessible from this point of view are "overriding"? Is it not obvious that when the verdicts of the "impersonal" point of view conflict with the verdicts of the "personal" point of view, it is the *latter* that have priority? After all, someone's personal verdicts are *her* verdicts. Why would she override them? Is it even possible to repudiate one's own point of view?

A similar argument begins with the observation that moral considerations are not the only kind. Why assume that what we value from the moral point of view is any better than what we value insofar as we are concerned to live a good life? The moral point of view is just one among others. So there is no good reason to privilege it over these others.

I hope to expose the confusion that underlies this way of understanding the challenge to minimally helpful careers, and thereby to expose the inadequacy of the two most popular responses to this challenge. To anticipate: the popular philosophical responses are wrong to suggest that doubts about the justifiability of minimally helpful careers presuppose a contestable point of view, constituted by contestable evaluative priorities. In fact, doubts about minimally helpful careers are, essentially, doubts about which points of view—which evaluative priorities—can be justified. I will try to show that there is no way to answer this question without ascending from the meta-evaluative depths to the norms above ground; we have no choice but to tackle the hard task of assessing the motley collection of rather mundane considerations for and against the choice of minimally helpful careers. When we do this we may be able to build a shallow, messy case for pursuing such careers. It is a mistake to think we can do better—a mistake about the nature of our reasons for action.

---

[11] See Wolf, "Moral Saints." Though R. Jay Wallace does not suggest that the "moral point of view" poses any challenge to minimally helpful careers, he does contrast the verdicts reached from this point of view with those an agent reaches insofar as she wants to live a "worthwhile and meaningful" life. See Wallace, "The Rightness of Acts and the Goodness of Lives," in Wallace, Philip Pettit, Scheffler, and Michael Smith, eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (New York: Oxford, 2004), pp. 385–411, on p. 395. There is, of course, nothing wrong with dividing reasons into kinds. As will soon become clear, the problem lies in suggesting that moral reasons are distinct from the kind of reasons that are accessible "from the agent's own point of view" (Wallace, p. 395), or that there can be no reason to give them priority over reasons of any other kind. We refer to the "the moral point of view" to single out our concern about how our actions affect others, or humanity, or the community of rational agents. To claim that verdicts reached from the moral point of view are not "overriding" is simply to claim that other considerations—considerations of other kinds—can, on occasion, outweigh moral considerations. This is a *normative* claim. It is, in short, precisely the sort of "shallower" alternative I have in mind in saying that I cannot solve my practical problem by appealing to deep facts about agency and values. And it is as open to challenge from the point of view of deliberative reflection as any other normative claim. (See also note 44.) This is, I think, at least part of what lies behind Joseph Raz's rejection of the assumption that deliberating about what to do essentially involves dividing the relevant considerations into kinds. That a consideration qualifies as "moral" does not in itself determine how it stacks up against other considerations. See Raz, *Engaging Reason* (New York: Oxford, 1999).

And what if, at the end of the day, we are still unsure? Under these circumstances, we have no choice but to make a choice. If my criticism is on target, then this necessity does not in itself justify any particular project; after all, we might be mistaken. Nonetheless, I will suggest, it does have implications for how we should regard a person's commitment to her projects—implications that should be familiar to anyone who recognizes that respect for finite rational agents necessarily includes an acknowledgment of their fallibility.

IN SEARCH OF A SOLUTION (I): THE SPECIAL NORMATIVE STATUS OF MY RELATIONSHIP TO MY OWN LIFE

Like most people, I believe that I have reason to help others in desperate need. But is this an *overriding* reason? More particularly, is my reason to help compelling enough to imply—*not* that I ought to sacrifice all but my own most basic needs (for food, shelter, clothing, and safety), and *not* that I should ignore my special obligations (for example, to my daughter, and to the students who expect me to show up tomorrow to teach)—but that I should at least begin to make a transition to a career that would be—*not* as helpful as any career could possibly be, but *more* helpful than the one I have chosen? Clearly, I do not care as much about the fate of most living things as I care about my own. But so what? Is this disproportionate self-concern justified? Does it justify my choice of a career that steers me away from helping?

Though it is clearly in my self-interest to make some accommodations to the interests of others, this accommodation need not involve going out of my way to help them. It certainly does not preclude my choosing to be a philosopher. Still, my belief that I have reason to help others is not grounded in self-interest. I care about their well-being independent of its effect on *my* well-being. More to the point, I believe that I would have reason to help them even if I did *not* care. So, I need to ask: How does this reason stack up against the reasons of self-interest and inclination?

Though I hate to admit it, the reason to help those in great need appears to be pretty compelling. Indeed, it seems to be more compelling than any reasons I could possibly have for pursuing a philosophical career: Could I really with any confidence stand up before a crowd of desperately needy people and tell them that though I probably could use my skills and energy in ways that were far more beneficial to them, philosophy has a claim to these skills and energy which is every bit as strong? How would I make my case?

Perhaps, however, this is not the right way to think about things. Perhaps, I should reject the assumption that I cannot justify my choice without establishing that philosophy is just as important as other people and their needs. After all, this way of thinking gives no weight to my special relation to myself. I am not a mere instrument for the improvement of others. I have my own life to live. I have my own projects. Surely, this fact, alone, has normative significance.

According to many philosophers, nothing could be more straightforward than the relationship between my commitments—and whatever other psychological states have the capacity to motivate me—and my reasons for action. What I have reason to do is, they claim, necessarily a function of what my preferences, proclivities, and goals make it possible for me to *regard* as a reason. A reason is a consideration that counts in favor of my doing something; and whether a consideration counts in favor of *my* doing something depends on what *I* am *disposed to favor.* On this conception of practical reasons, solving my problem is as easy as getting clear about my own dispositions. Self-understanding is the key to what I have reason to do.[12]

No sooner do I rehearse this advice, however, then I see how useless it is. Indeed, my uncertainty about the justifiability of my career choice is compelling evidence against the view that what I have reason to do depends on what I am disposed to find reason-giving. It is very clear to me that I would *rather* devote my time, energy, and talent to philosophy than to any more helpful projects; I am strongly *inclined* to regard the imperatives of a philosophical life as sufficient reason not to do more to help others in need. Yet it is equally clear to me that this settles nothing. My question is: What should I *make of* these inclinations and preferences? What is the normative status of the "elements" in my "motivational set"?[13]

---

[12] Bernard Williams defends this roughly Humean view in "Internal and External Reasons," in *Moral Luck* (New York: Cambridge, 1981), pp. 101–13. Frankfurt has also devoted many papers to arguing that what we have reason to do ultimately depends on what we care about. "The most basic and essential question for a person to raise concerning the conduct of his life," he argues, "cannot be the *normative* question of how he *should* live. That question can sensibly be asked only on the basis of a prior answer to the *factual* question of what he actually *does* care about" (Frankfurt, "The Question: How Should We Live?" in *The Reasons of Love* (Princeton: University Press, 2004), pp. 3–32, on p. 26). See also "The Importance of What We Care About," and "Rationality and the Unthinkable," in *The Importance of What We Care About*, pp. 80–94 and 177–90; "On Caring," in *Necessity, Volition, and Love* (New York: Cambridge, 1999), pp. 155–80; and Replies to Watson, Wolf, and Herman, in *Contours of Agency*, pp. 160–64, 245–52, and 275–78.

This is a perfectly intelligible question. Indeed, my commitment to philosophy would not be the strong motivating force that it is if I were convinced that I had no good reason to sustain it. If my commitment remained strong, despite my belief that it was unjustified, I would be irrational; and if I nonetheless treated it as a guide to my life, I would be weak-willed. The mere fact that I happen to be drawn to the philosophy project is no more intrinsically reason-giving than any other fact. It *reflects* my belief that philosophy is a project worth pursuing. So reminding myself that this is, indeed, my belief does nothing to reassure me that I am not (and could not possibly be) mistaken.

Still, even if I cannot justify my career choice by appealing to the motivational force of my commitment to philosophy, is there not *some* connection between the fact that this commitment is *mine* and the fact that I have sufficient reason to maintain it? Surely, it matters that the commitment is a very central aspect of my identity. To abandon this commitment would be to repudiate who I am. If this would not exactly involve my *contradicting* myself, would it not be every bit as unreasonable? Would it not be irrational? How can *I* have reason to do something that is incompatible with my *being me*?

Sartre reminds us, of course, that a person can change radically without ceasing to be the person she is; someone who declares "I am what I am" in order to justify her actions is guilty of bad faith. But surely it is more complicated than this. For, Sartre's insistence to the contrary notwithstanding, we cannot simply pick our reasons and values. We cannot will them into existence. We need to discover them; and we cannot do this without relying on the commitments that structure our reasoning. Unless I am *something*, there is nothing about me that will lead me to take any fact as the basis of an inference; I will discover no reason to do anything at all. Does this not suggest that "I am what I am" is really, not a poor excuse, but a genuine justification?

If I had no projects, no commitments of any kind, my identity would, it seems, be so thin that I would lack the capacity to choose from among the many options that confront me; none of my options would appear to be more choiceworthy than any other. This suggests that I have very good reason not to abandon my projects, even if they are not particularly helpful to others: my projects determine what can count for me as a reason. How, then, can I be mistaken to treat my philosophy project as a basis for normative conclusions?

---

[13] Williams, "Internal and External Reasons."

Many philosophers have attempted to develop these vague thoughts into a defense of the choice to lead a relatively unhelpful life, or at least to keep leading such a life if one has already chosen to do so. (The ultimate target of the arguments is much broader: they challenge a wide range of moral theories and the broad assumptions underlying these theories; in so doing, they imply, among other things, that most of us do not have overriding reason to choose more helpful projects.) Importantly for my purposes, these arguments do not presuppose that the only available alternatives are lives of supreme self-sacrifice, or of single-minded pursuit of "moral interests." If the arguments work, they imply that I do not have, and never had, an overriding reason to commit myself to the sort of projects that are more likely to help others in need, even though I could have done so (and could still do so) without abandoning my family, sacrificing my friendships, or giving up my "free time" and the various activities in which I like to engage during that time.

Bernard Williams's argument is probably the best known.[14] But similar points have been made by Christine Korsgaard,[15] Harry Frankfurt,[16] and Samuel Scheffler,[17] among others.[18] In considering the defense these philosophers offer me, I will downplay the disagreements among them.

All versions of this defense begin with the observation that whatever formal principles of rationality there may be, they do not provide me with grounds for choosing among all of the otherwise acceptable possibilities.[19] The next move is to ask what conditions must be satisfied if I am to make the further discriminations. The highly plausible answer is that I need to care about certain things, to be committed to them, to identify myself with certain dispositions and certain roles. Only under these circumstances will I be disposed to regard certain facts as more compelling reasons than others. And only if I am so disposed will I be capable of making up my mind about what to do. If we think of a personal "ground project"[20] as a set of identity-conferring, closely related, mutually reinforcing, somewhat general commitments and concerns, like those tied to one's professional role, or one's role as a parent, then to abandon one's ground projects—or even to act in a way that is incompatible with them—would leave one without the ability to justify most of one's choices.[21] In short, having projects is a necessary condition for the possibility of having sufficient "material" for deliberation. It is our ground projects—our "practical identities"[22]—that structure, or govern, our deliberations by determining the normative significance (for us) of the various facts we consider. We thus have very good reason *not* to abandon our projects: our ground projects, in particular, determine what we have reason to do.

---

[14] See Williams, "Persons, Character, and Morality," and "Moral Luck," in *Moral Luck*, pp. 1–19 and 20–39; "Integrity," in *A Critique of Utilitarianism*, in J.J.C. Smart and Williams, *Utilitarianism: For and Against* (New York: Cambridge, 1973), pp. 108–18; and "The Makropolis Case," in *Problems of the Self* (New York: Cambridge, 1973), pp. 82–100.

[15] See Korsgaard, *Sources of Normativity* (New York: Cambridge, 1996), pp. 90–130.

[16] See Frankfurt, "On the Necessity of Ideals," in *Necessity, Volition, and Love*, pp. 108–16. See also Frankfurt, Replies to Velleman and Watson, in *Contours of Agency*, pp. 124–28 and 160–64.

[17] See Scheffler, *The Rejection of Consequentialism*.

[18] Thomas Nagel has also made an important contribution to the discussion of the relationship between reasons and "points of view." See Nagel, "Subjective and Objective," in *Mortal Questions* (New York: Cambridge, 1979), pp. 196–213, and *The View from Nowhere* (New York: Oxford, 1986).

[19] Notice that on this assumption, if we have reason to "maximize happiness," this is not because we would be *irrational* to will to do otherwise.

The same point can be put somewhat differently. Since our identity is tied up with our ground projects—since they would not be our ground projects if we were not "identified with" them[23]—to act contrary to the constraints they impose on us is, in effect, to contradict ourselves. We cannot abandon our ground projects without abandoning ourselves. And to be at odds with oneself in this way is a paradigm case of irrationality.

---

[20] See Williams, "Persons, Character, and Morality."

[21] A person, Williams explains, "is identified with his actions as flowing from projects and attitudes which in some cases he takes seriously at the deepest level.... [*H*] *is* actions and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified" (Williams, *A Critique of Utilitarianism*, pp. 116–17). Similarly, Frankfurt argues that someone with no interests and preferences "is so vacant of identifiable and stable volitional tendencies and constraints that he cannot deliberate or make decisions in any conscientious way. If he nonetheless does remain in some way capable of choice, the decisions and choices he makes will be altogether arbitrary. They cannot possess authentically personal significance or authority, for his will has no determinate character" (Frankfurt, "On the Necessity of Ideals," p. 110).

[22] Your "practical identity," Korsgaard explains, is the "description under which you value yourself, a description under which you find your life to be worth undertaking.... You are a human being, a woman, or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain profession, someone's lover or friend, and so on. And all of these identities give rise to reasons and obligations..." (*Sources of Normativity*, p. 101).

[23] As Scheffler explains, "If I value my projects non-instrumentally, then I will see them as providing me with reasons for action in a way that the projects of other people do not, and in a way that other activities now open to me do not." "To value my own projects just is, among other things, to see them as providing me with reasons for action in a way that other activities and other people's projects do not....

The implications for my own case are clear: if one is committed to a philosophical career, then one has a very good reason to continue on this path. Insofar as I am a philosopher, I necessarily evaluate my options from the point of view of a philosopher; I am governed by the imperatives of a philosophical life. Moreover, this point of view does not violate any principles of rationality. It is compatible with the defining commitments of rational agency; I do not contradict myself when I treat the reasons of philosophy as reasons not to be more helpful.

Since Reason itself does not speak against the philosophical life, there is, the argument continues, no independent point of view from which I might settle the question of whether I *really do* have the reasons I appear to have when I take my philosophical career for granted—no point of view other than those that are constituted by other, equally contingent, projects. There are, of course, points of view from which the reasons of a philosopher are invisible. And there are other points of view from which these reasons lack the weight that I attribute to them. But I have no (overriding) reason to adopt these points of view myself. Indeed, I have good reason not to: my identity as a philosopher supplements the requirements of pure reason not only because it is compatible with these requirements but also because it imposes requirements itself.[24]

I might not have chosen to be a philosopher; I might have identified with a different role, embraced a different identity. But I chose the philosophy project. And so, the argument tells me, I cannot coherently regard my identity as a philosopher as a contingent fact that I

[I]f I do not see myself as having any more reason to pursue my own projects and goals than I do to engage in other comparable activities, then it no longer makes sense to say that I value those projects and goals. Indeed, it no longer makes sense to describe them as my projects and goals at all" (Scheffler, "Projects, Relationships, and Reasons," in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, pp. 247–69, on pp. 254–55).

Frankfurt has defended the normative importance of "identifying with" one's motives in an influential series of papers. See, for example, "Freedom of the Will and the Concept of a Person," and "Identification and Wholeheartedness," in *The Importance of What We Care About*, pp. 11–25 and 159–76.

[24] For an interesting discussion of the conflict that can arise between the requirements grounded in one's particular practical identity and the requirements grounded in one's identity as a rational agent, see Gerald Cohen's commentary on Korsgaard's Tanner Lectures, and Korsgaard's reply (Cohen, "Reason, Humanity, and the Moral Law," and Korsgaard, "Reply," in *Sources of Normativity*, pp. 183–84 and 254–58).

might change whenever I want. Changing careers is not a live option for me—or at least it has a lot less life in it than I have been assuming.

Again, to insist otherwise is to deplete my supply of reasons for action. Worse still, it is to acknowledge this fact. No wonder my search for self-justification seems to involve so much fruitless thrashing around! My attempt to defend myself is, apparently, the hopeless enterprise of someone who necessarily identifies with the concerns and commitments of a philosopher despite being alienated from this very identity!

I believe that I ought to avoid fruitless thrashing around. One way to achieve this end would be to refrain from the self-alienating move: I could accept the fact that I am a philosopher and that, accordingly, I have the reasons of a philosopher. But surely, this is the lazy, cowardly way out. It is the way of bad faith: I am what I am, end of story. Even if my reasons do depend on my identity to an extent that Sartre overlooked, my identity is not self-justifying. So the fact that I am who I am does not justify my disinclination to change. I do not contradict myself whenever I challenge the dispositions that constitute even a very central aspect of my identity; at most, I contradict some of the assumptions I make, given that I have these dispositions.[25]

To be sure, there is a sense in which, as Korsgaard puts it, to "violate" "the conceptions of ourselves that are most important to us" is to "no longer be who [we] are."[26] But to suggest, as she does, that "when an action cannot be performed without loss of some fundamental part of one's identity," then one "could just as well be dead"[27] is to obscure the difference between the sense in which I am the subject of my projects and the sense in which I am a living thing. I do not mean to deny that there could be conditions under which I am alive and yet could just as well be dead. Loss of my capacity to reason and perhaps of big chunks of what we vaguely call my "personality" are compelling candidates. My point is that the loss of my projects is different: even if in an obvious sense they are a very important part of my identity, even if I cannot help "identifying with" some of them; continuing to pursue them is not a necessary condition for continuing to be me.

---

[25] For an interesting critique of the role that identity plays in Frankfurt's account of practical reasons and self-governed action, see David Velleman, "Identity and Identification," in *Contours of Agency*, pp. 91–123. For Frankfurt's reply, see pp. 124–28. (See also note 26.) Velleman does not deny that there are connections between who we are and what we have reason to do. His point is just that we may have reason to change who we are. He says, for example, that "the Mafioso" discussed by Korsgaard and Cohen (see note 24) has reasons "to revoke his commitment to that more particular identity, which turns out to conflict with his underlying identity as a human being; and so they are reasons for him to become someone who no longer has reasons for committing murder. The mobster is irrational to commit murder, not because he doesn't have reasons for committing such an act, but rather because he has reasons against being the sort of person who has those reasons" (Velleman, "Willing the Law," in *Self to Self* (New York: Cambridge, 2006), pp. 284–311, on p. 306). I myself do not object to this way of putting things as long as it is equivalent to saying that the Mafioso has stronger, better reasons for *not* committing murder.

[26] Korsgaard, *Sources of Normativity*, p. 102. Similarly, according to Frankfurt, "Agamemnon at Aulis is destroyed by an inescapable conflict between two equally defining elements in his own nature: his love for his daughter and his love for the army he commands.... When he is forced to sacrifice one of these, he is thereby forced to betray himself. Rarely, if ever, do tragedies of this sort have sequels. Since the volitional unity of the tragic hero has been irreparably ruptured, there is a sense in which the person he had been no longer exists" (Frankfurt, "Autonomy, Necessity, and Love," in *Necessity, Volition, and Love*, pp. 129–41, see p. 139).

It is important to be clear about this. I am willing to concede that certain projects may be as essential to my personal identity as certain activities are essential to the biological identity of some animals. But that is just the point: for any animal that is capable of practical reasoning, the notion of an identity-determining activity is problematic. If a wolf could ask itself whether sheep really are to-be-eaten, it could not settle the matter by noting that it is, after all, a wolf, and wolves eat sheep. That it is a wolf is *a* reason in favor of its eating sheep. But if it could survive in other ways—and especially if it could thrive, or at least believes that this might well be possible—then it could not answer its own question without first asking itself whether it has good reason to *be* a sheep-eater—good reason, that is, to be a wolf.[28]

It sounds like a joke, but apparently it needs to be said: living is essential to my being who I am in a way that "doing philosophy" is not. I am *nothing* if I am not alive. So the question of whether to keep living is not a question of which identity I should embrace. It is a question about whether to be anything at all. This, too, is an intelligible question. But, Williams's suggestion to the contrary notwithstanding, I do not need to be invested in any projects in order to give an affirmative answer. It is enough that in having been born, I have been given the task of living. I need a good reason to quit—not to stay on the job.

If I had no projects, I would not, for this reason alone, be stuck with a mere "drive" to stay alive;[29] regardless of whether any such drive persisted, I would take myself to have reason to encourage it. My implicit recognition of the connection between being alive and needing a good reason not to keep living is what underlies the ease with which I invest myself in my various projects. In short, at least in my own case, Williams has things backwards. My conviction that I have sufficient reason to live does not depend on my having certain basic commitments, or projects. To the contrary, I would not bother choosing among various projects if I did not think I had sufficient reason to live.[30]

---

[27] Korsgaard, *Sources of Normativity*, p. 102. As Korsgaard elsewhere explains, "Unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another—and with it, your grip on yourself as having any reason to live and act at all" (*Sources of Normativity*, pp. 120–21).

[28] For an extended discussion of what would follow if a wolf could reason about the pros and cons of serving the interests of the pack, see John McDowell, "Two Sorts of Naturalism," in Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn, eds., *Virtues and Reasons: Philippa Foot and Moral Theory* (New York: Oxford, 1995), pp. 149–79, on pp. 151–55.

[29] Williams, "The Makropolis Case," p. 86.

To violate an imperative intrinsic to my projects is, then, neither to contradict myself nor to abandon myself.[31] It certainly does not cause me to cease to be who I am. To think otherwise is to mistake a deep diagnosis of why it is so difficult for someone to question the normative status of her own projects for a proof that she has made a conceptual mistake if she wonders whether she really ought to pursue these projects. The diagnosis is deep in the sense that it does not attribute the difficulty to a mere psychological flaw: it shows how our tendency to resist challenges to our projects is tied to the structure of reasons for action, and so has a certain justification. But this does not mean that such challenges are irrational, or even unreasonable. For even if our reasons necessarily depend on our projects, this is compatible with the fact that no particular project is itself necessary. Though their being our projects necessarily involves our not regarding them as optional, we have considerable say over whether they are our projects. We have the power to reshape them, and even to replace them.[32] So it is hard to see how we can justify closing the question as to whether this is just what we have good reason to do.

---

[30] Notice that this point has the form of a transcendental argument: it is not just that I do, in fact, take my being alive to be a reason to keep on living, but that I would not take myself to have reason to do whatever is involved in "living a life" if I did not believe that life was (for me) worth living. For a similar critique of Williams, see Frankfurt, Tanner Lectures, 2004 (unpublished manuscript), and "The Necessities of Love," Part II of "On Caring, " in *Necessity, Volition, and Love*, pp. 167–80, on p. 172. Whereas Frankfurt argues that our *love* of life provides us with reason enough to stay alive, my point is that—whether we love life or not—we generally take ourselves to be justified in staying alive, independent of any other commitments, and this assumption makes sense of our other assumptions about what we have reason to do.

[31] In short, it is a mistake to think of the commitments one makes to oneself as analogous to the commitments one makes to others. Though one cannot commit to doing *X* without intending to close the question as to whether one will do *X*, this is not necessarily a sufficient reason *not* to reopen the question. One can change one's priorities and goals without betraying oneself. This is why there is, as Scheffler notes, "a fundamental difference" between the "wrong-making" reasons of parenthood and friendship and "the project-dependent reasons we may have to devote resources to music or golf. Those project-dependent reasons differ from relationship-dependent reasons not in being 'even more clearly' optional in character, but in being reasons we can neglect without doing anything wrong" (Scheffler, "Projects, Relationships, and Reasons," p. 264). For an extended discussion of the constraints our intentions impose on our reasoning, see Michael Bratman, *Intentions, Plans, and Practical Reason* (Cambridge: Harvard, 1987).

Of course, self-criticism necessarily takes place from some point of view. Necessarily (and trivially), this point of view is my own. (It is, if you like, my own "personal" point of view.) Yet it is also a point of view from which I challenge my own take on things. So, one might ask: Where does it get its authority? Why should I privilege it over any other (possibly more reassuring) alternative?

The first answer is that one can challenge one's take on things without thereby abandoning it, and without presupposing that any other set of substantive normative assumptions is superior; one can acknowledge that one's practical commitments may be unjustified without thereby relying on an alternative authority whose status is suspect from the point of view of the commitments themselves. Just as one can appreciate that one's empirical beliefs may be false even though one has no grounds for challenging any belief in particular, so too, one can appreciate that a given normative claim may be false without offering any particular reason for rejecting it.

The second, and more direct, answer is that the point of view from which I question my own commitments has no more and no less authority than the point of view constituted by the commitments themselves. And this is because the one point of view is not really external to the other. My commitment to being guided by reason is an integral part of every point of view from which I evaluate my options, no matter how idiosyncratic this point of view happens to be. So, if I discover (or at least think I discover) a reason to alter my commitments, there is, necessarily, no point of view from which I am indifferent to this discovery.

If there is any antagonism between, on the one hand, what one might call the more "impersonal" point of view from which one regards one's (personal) projects as optional and, on the other hand, the points of view of the projects themselves, this antagonism is due to the fact that to take the reason-giving status of one's projects for granted is precisely *not* to regard them as optional. But since, whatever my personal commitments may be, I am always committed to acting for reasons, even this antagonism is distorted by talk of the personal versus the impersonal. Again, we have no choice but to regard our projects as constraints on what we have reason to do. But again, we respect these constraints only because we believe that they are legitimate guides to what we have reason to do. The basis for challenging our projects thus has its origin in the normative status of the projects themselves.

---

[32] The same point can be put in a slightly different way. Even though it is not possible for me to regard my projects as those of just *someone*, it is perfectly intelligible for me to ask: "What should *my* projects be?"

In sum, my ground projects are not just very important to me. They play an essential role in structuring/governing my deliberations about what to do. In order for them to play this role—in order for them to be truly my projects—I cannot treat them as optional. At the same time, however, I know that I will still have reasons for action if I have to flee my country and "start all over again" flipping burgers; I will still have reason to live.[33] Knowing this, I can wonder about the desirability of starting over. And to this end, I can wonder about the justifiability of *not* starting over. I can ask: Am I justified in pursuing a philosophical career?

Given that being a philosopher is such an important aspect of my identity, it requires quite a struggle to take this question seriously. And the temptation to dismiss it is strengthened by the fact that I seem to have no principled way of answering it. I seem to have reasons of such different kinds, and no systematic way of adjudicating among them. This encourages me to treat the question as closed, and to carry on as before. Yet treating it as closed does not suffice to close it. The question remains open.

Thrashing over such questions is, I know, a sign of privilege. Not only does it require a fair amount of leisure time, but an agent must have genuine options. Very few human beings satisfy these conditions. Most are too busy surviving from one day to the next to worry about what sort of life they should lead. And most of the rest have little choice but to accept the roles their society assigns them: the fact that they have a certain gender, belong to a certain social group, and so on, involves normative presumptions that are at least as strong as the presumption in favor of staying alive. Because under these social conditions, it is usually quite clear what one has reason to do, there are some who promote highly "traditional" societies as an antidote to thrashing.[34]

---

[33] And—it is worth stressing again—I will still be a well-functioning, flourishing agent. Pressing the apparent analogy between my relationship to myself and my relationship to my child, Richard Miller offers a normative defense of neglecting others in need: I have a special responsibility for my own care. See Miller, "Beneficence, Duty, and Distance," *Philosophy and Public Affairs*, xxxii, 4 (2004): 357–83, on p. 370. Even if the analogy were a good one, however, it would not imply that I am justified in choosing a relatively unhelpful career. For, as I have already observed, this choice is not one about which I can in good conscience declare that it "imposes a significant risk of [significantly] worsening [my] life" (p. 370).

[34] See, for example, Alasdair MacIntyre, *After Virtue* (Notre Dame, IN: University Press, 1981).

This solution cannot help *me*, however. And in any case, enthusiasm for eliminating the contingency of our own personal projects is, it seems to me, no less a sign of laziness and cowardice than is the refusal to acknowledge the vertiginous implications of this contingency.

IN SEARCH OF A SOLUTION (II): THE VALUE OF NONHELPFUL PROJECTS

As I have just indicated, the confusion caused by my having *so little* to lean on in sorting out my options is but the flip side of the confusion caused by my having *so many* different kinds of reasons. When the question of whether I should switch careers is open to me, I see that I cannot arrive at an answer without "weighing" my reason to help others against whatever reasons support and are reinforced by my personal concerns and commitments. Yet there seems to be no scale for the job. So I seem to have no choice but to look around for still more relevant considerations—more reasons, more values—anything that, in relation to everything else, might prove decisive.

Notice that it will not help at this point to observe that I am not rationally compelled to use a "moral" scale; for my difficulty does not rest on assuming otherwise. The point of view from which I confront the career question is, necessarily, *my* point of view; and it is, necessarily, the point of view of someone who is moved by reasons—whatever reasons I am able to comprehend. If there is a distinct point of view we might attribute to "morality" —the so-called "moral point of view" —then I am interested in the verdicts reached from this point of view only because, like the verdicts reached by Tom, Dick, or Harry, they might well be justified.[35]

Presupposing nothing, in particular, about morality, then, I ask myself: Which other considerations are relevant to assessing the relative significance of the fact that I have reason to help others and the fact that I have the reasons of a philosopher?[36] In search of such considerations, I naturally turn to the value of my chosen career: a philosophical life is a valuable life. I know, of course, that the same can be said, with equal justification, of other lives that hold out the promise of allowing me to be more helpful to others. But if, as seems to be the case, the value of helping and the value of doing philosophy are of irreducibly different kinds, then how can I possibly evaluate their relative worth? Have I not come as far as reason can possibly take me? Or, if there is still a relevant consideration left over, is it not the fact that variety itself is valuable? Does this not count in favor of my decision? After all, if everyone were to abandon the less helpful careers, the variety of valuable human lives would be greatly diminished; the world would be a much drabber place.[37]

---

[35] This represents a response to those who suggest that I can somehow remove the force of my question by identifying it as a moral question, and then contrasting the "moral point of view" with the point of view from which the chief concern is "eudaimonia," or "perfection." (See also note 11.) As I indicate below, insofar as the defense of nonmoral considerations is not simply a *normative* one, it has essentially the same structure as the one we just reviewed—and it fails for essentially the same reason.

[36] As many have argued in discussing animal rights, it does seem that my reason to help is relatively weak when those whose lives are at stake have no stake in these lives—when they could not possibly care less whether they ceased to exist. But this is not relevant to the problem I am grappling with here. After all, I am not worried about neglecting the grass and the mosquitoes. It is other human beings whose claim to my help I want to find reason to override.

Whereas the previous argument appealed to the connection between a person's identity-defining commitments and her reasons, this argument appeals to the alleged incommensurability of values. Whereas the first argument suggested that my choice is justified by the fact that I am who I am (and that no one but me can initiate my actions), the second suggests that my choice is justified by the fact that nothing (nothing about me, or about anything else) rationally compels me to choose one way or another. Despite these important differences, however, the two arguments share an important element: each justifies my career choice by appealing to the apparent fact that there is no privileged point of view from which to evaluate my disposition to make this choice. According to the first argument, this frees me (or, if you like, forces me) to treat my dispositions to choose (at least those dispositions that are most central to my take on things) as a source of reasons. According to the second argument, it frees me to pick any criterion of choice that is compatible with at least some true evaluative judgments. As Susan Wolf explains, "If there is a justification for the institution of [relatively unhelpful careers], it is one which rests on the *decision not* to...justify every activity against morally beneficial alternatives."[38]

Unfortunately, the fact that there is no single, privileged point of view from which to determine the relative merit of every valuable project no more justifies relying on a standard of appraisal that supports the choice of unhelpful careers than it justifies relying on personal commitments that support this choice. Even if values are irreducibly plural, this does not license me to just "decide" which considerations are relevant to the justifiability of my choice; the fact that different projects are good in different ways does not give me a sufficient reason to pick whichever valuable career I happen to like. Nor is such justification to be found in the value of variety.

---

[37] See Wolf, "Moral Saints."

[38] "Moral Saints," p. 422; the emphasis is mine.

In short, the variety of values does not prevent me from making a reasoned choice between options with different values; and the value of variety, in particular, is often trumped by other valuable things. If I am tempted to think otherwise, this is, I suspect, because I am inclined to wishful thinking. Indeed, the appeal to how drab the world would be without minimally helpful projects bears an uncomfortable resemblance to certain theodicies whose persuasive force was (and still is) largely a function of the fact that finding them convincing seems to be the only way to protect certain dearly held beliefs and ways of life. Just as the value of God's creation does not imply that the world's evils are justified, so too, whatever values are associated with unhelpful careers do not necessarily constitute a justification for neglecting other human beings in need; nor does the fact that these values are irreducibly distinct from the values that structure far more helpful lives.

There are occasions on which my choices between competing valuable options seem arbitrary: I seem to pick one of the options simply because I cannot pursue more than one. On many other occasions, however, I seem to be able to determine the relative merit of different valuable alternatives.[39] Thus even though the value of a human life and the value of art are irreducibly different kinds, I take it to be obvious that I have a good reason not to kill someone who stands in the way of a large allocation to the National Endowment for the Arts—even if I could get away with it.

This example calls attention to the familiar difficulty of finding a defense of minimally helpful projects that is not, at the same time, a defense of harmful projects.[40] A paradigmatically harmful project could, it seems, be no less essential to someone's identity than a minimally helpful project. So we seem to need an account of why identity-determining commitments are adequate justification in some cases, but not others. Similarly, if we are to justify minimally helpful projects by appeal to the fact that there are many different kinds of value (and that variety itself is a value of a certain kind), we must explain why harmful projects whose harms are a means to valuable ends do not qualify as one kind of "valuable project"; or, if they do qualify, we must explain why only some valuable projects are justified by the fact that their value cannot be reduced to the values that count against them.

---

[39] As David Wiggins notes, "often, in a given case, we *can* find a way to collate and arbitrate between the demands of... impartiality, benevolence, mercy, due process, and so on. There may, however, be no general method of doing this" (Wiggins, "Incommensurability: Four Proposals," in Ruth Chang, ed., *Incommensurability, Incomparability, and Practical Reason* (Cambridge: Harvard, 1997), pp. 52–66, on p. 58). According to Wiggins, "we have to make our choice in the light of our overall practical conception of how to be or how to live a life (both here and in general). We deploy these conceptions even as the variety of the contingencies that we actually confront constantly shapes or reshapes the conceptions themselves (a two-way flow). It will be no wonder if choice (as now described) is the exercise of an irreducibly practical knowledge that can never be exhaustively transposed into any finite set of objectives that admit of finite expression" (p. 59). Interestingly, Wolf says something similar: "both in our philosophizing and in our lives, we must be willing to raise normative questions from a perspective that is unattached to a commitment to any particular well-ordered system of values" ("Moral Saints," p. 439). (See also note 44.)

[40] There is, of course, voluminous literature on the apparent moral difference between harming and not helping. For a thorough critique of the alleged distinction, see Kagan, *The Limits of Morality*.

Again, the prospects of finding such arguments do not seem especially good. For again, there is no obvious connection between the incommensurability of values and the license to just pick which values will guide one's decisions. Nor does the mere fact that each of our options is valuable in irreducibly different ways license us to allow ourselves to give priority to whichever values happen to have the greatest power to motivate us.

If the world's entire collection of poetry (or all that remained of nineteenth and twentieth century paintings after the last nuclear war) were housed in a burning building, and if I could save some of the precious volumes (canvasses) by neglecting a child trapped there with them, I would, I believe, have overriding reason to save the child. The loss would be great; I am prepared to admit that it would be greater than the loss of the child. But this does not seem to give me a good enough reason to leave her to burn. I have no satisfactory story to tell about why this is so. I know that some will think I am mistaken.[41] My point, however, is that however distinct the values at issue, I do seem to manage to "weigh" them against one another. What is more, when I weigh them, I often discover that I have sufficient reason to forfeit variety, and indeed just about every other value that is distinct from the values at stake in preserving human life. This means that I cannot justify choosing a relatively unhelpful career by simply calling attention to the fact that this type of life is very valuable indeed, nor even that it has a value that is irreducible to any other. I cannot solve my problem by simply calling attention to the plurality of values.

STILL SEARCHING

It is, George Eliot reminds us, "a common trick of desire" to "avail itself of any irrelevant skepticism, finding larger room for itself in all uncertainty about effects, in every obscurity that looks like the absence of law."[42] Every human being has desires. So every human being is vulnerable to this trick.

---

[41] This has been the reaction of more than one philosopher upon hearing or reading this paper.

Under the pressure of various desires, it is tempting to think that the very nature of reasons and values limits the extent to which it is reasonable to expect one human being to help another; it is tempting to think that the choice to pursue a relatively unhelpful career—the choice to pursue a career regardless of how beneficial it may be to others, especially others in great need—can be justified by the nature of rational agency or the plurality of values. If I am right, however, the two most compelling philosophical arguments along these lines are no good. If we are inclined to endorse them, this is a symptom of our susceptibility to bad faith, and to theodicy-like rationalizations. If some of us really do have good enough reasons to devote ourselves to relatively unhelpful projects, these reasons are far less deep than many philosophers have supposed. To find these reasons, we must turn our attention to the shallow places where considerations pro and con multiply beyond reckoning.

It is important to see how this conclusion differs from another familiar response to the suggestion that we are justified in neglecting human beings in need. In *The Limits of Morality*, Shelly Kagan considers various alleged reasons for thinking that the value of certain valuable things and relations is inaccessible from the "objective standpoint" we occupy when we do not privilege our own interests and concerns over others. If this were true, he says, and if these values were accessible from the "subjective standpoint," then the subjective standpoint would have independent authority; we would have good reason to give our own interests—and commitments and projects—more weight than we give to the interests, commitments, and projects of others.[43]

But this is confused. If something's value is relative to a given standpoint, then this value cannot be used to defend the authority, or value, of the standpoint itself. If various human activities are valuable, then the fact that they are valuable is not a fact about what follows from being disposed to value things in a certain way. No wonder, then, that Kagan has little difficulty establishing that we need not occupy the "subjective standpoint" in order to value these valuable things!

---

[42] Eliot, *Middlemarch* (New York: Penguin, 1994), p. 705.

[43] In each case," Kagan writes, "the moderate has indicated something of value (for example, the obtaining or possession of some good or relation). He has claimed, in effect, that the objective standpoint is unable to offer an account of these values, and so it is inappropriate for an agent to attempt to act in accordance with a totally objective point of view. Since it is only a (partially) subjective point of view which can adequately incorporate the values, there is reason to maintain something of the natural bias which is involved in the subjective standpoint" (*The Limits of Morality*, pp. 369–70).

Whereas the arguments Kagan considers appeal to substantive values to justify neglecting the needy, the arguments I have considered operate on the *meta-evaluative* plane. In effect, they are transcendental arguments in defense of a relativism that is supposed to be free of the familiar problems with relativism precisely because it is transcendentally justified.[44]

Insofar as the debate over "prerogatives" is a debate over whether a given activity or relationship or way of life is really valuable, and over how its value is to be weighed against other valuable things, the debate is a *normative* one; and insofar as it is a normative debate, it is—as Kagan, in effect, assumes—a debate over what we have reason to do. This, I have tried to show, is what is really at stake in arguments about whether we ought to do more to help those in great need. All the talk about standpoints, all the talk about the necessary conditions of rational agency and the plurality of values, merely distracts us from the task we face in trying to figure out how to live our lives.

---

[44] Wolf's argument appears to be a hybrid. She alludes to substantive values which she thinks "morality" cannot acknowledge. But she also suggests that, since there is no "well-ordered system of values" more fundamental than the values accessible from either the "point of view of individual perfection" or the "moral point of view," there can be no argument for favoring the verdicts reached from the one point of view over the verdicts reached from the other. In contrast, Wallace suggests that it is possible to "[show] how compliance with moral norms can contribute directly to making our lives worthwhile or meaningful, so that we ourselves live better to the extent we achieve such compliance" (Wallace, "The Rightness of Acts," p. 406).

In arguing that the normative status of morality depends on this possibility, Wallace also combines, to an even greater extent than does Wolf, normative appeals and appeals to the nature of agency and values. According to Wallace, the "eudaimonistic standpoint" is the standpoint from which we are interested in "whether engaging in [certain] pursuits is *worthwhile*, something that makes our own lives choice-worthy as human lives. [From this standpoint,] we...ask ourselves whether our life is made meaningful or worth living as a result of our being engaged in these pursuits" (pp. 395–96). Wallace asks whether "morality might be rationally deficient by comparison with the personal relationships and endeavours around which we structure our lives" (p. 395). This sounds like a normative question. But as we read further, the concern seems to be whether moral demands are compatible with the necessary conditions of rational agency. Eudaimonistic considerations have priority over moral considerations, Wallace urges, in the sense that morality owes its normative status to the contribution it makes to a meaningful life. Rather than directly defend this normative assumption, however, he takes it to follow from the fact that the eudaimonistic standpoint is "the agent's point of view" (p. 395). There might seem to be nothing more to this fact than the fact that our interest in discovering what is "meaningful," "significant," or "worthwhile" is an interest in discovering what "we have reason to engage in" (p. 396). But, of course, a commitment to discovering reasons has no particular normative implications. Wallace thus insists that he has something more determinate in mind (p. 396); the eudaimonistic standpoint is one standpoint among others. If, however, the eudaimonistic standpoint is just one standpoint an agent can take up, then it is not the only standpoint from which considerations can have normative significance "for her." So, again, she needs some reason to give this standpoint priority over others. That is, she needs a reason for giving special weight to whatever considerations have eudaimonistic value. "I would have to work hard not to find this pursuit boring," she thinks to herself. "But what," she naturally wonders, "does this imply? Are there any considerations that count in favor of the pursuit? If there are, then how do they stack up against my desire to avoid pursuits that tend to bore me?"

Once we leave the distractions behind and get down to business, we have no choice, I have said, but to try to sort through the relative importance of various valuable things and relations. I have argued elsewhere that this is, essentially, a matter of identifying various facts whose normative status depends, in turn, on still other facts; and that there are no facts that necessarily bring an end to this process.[45] For my purposes here, however, the important point is the negative one: no normative assumptions, no matter how dear, are beyond challenge. This does not mean, of course, that we can never make normative assumptions. Clearly, we must do so in order to get on with things; our very sanity depends on our ability to circumscribe our normative inquiries. Nonetheless, it is surely appropriate—and even desirable—for human beings to ask themselves: How should I structure my life? And in order to answer this question, we have no choice but to ask others too: Can I justify my present identity? How important is it, really, that I have already committed myself to a certain way of life? How important is it whether this way of life is useful? How do the reasons for maximizing value stack up against the reasons for acknowledging values in other ways? How do the needs and interests of others stack up against my own needs and interests? What, exactly, are my own needs and interests?

In grappling with these normative questions, a person may well conclude that a community of human beings has good reason not to compel its members to pursue more helpful projects. But it is important to see that this conclusion does not itself put the questions to rest. There may well be good reasons for my society not to punish sexual relations outside marriage. Yet it hardly follows that I have good reason to commit adultery. There may be good reasons to allow parents to teach their children that certain ethnic groups ought to be pushed into the sea—good reasons not to remove these children from their homes, even if this means that their minds will be filled with poisonous falsehoods. But this does not mean that I have good reason to teach my daughter that the Blacks, the Jews, the Gays, and the Liberals are a threat to the American way of life.[46] The normative issue is not: Would a just society use its power(s) to discourage its members from pursuing minimally helpful careers? It is, rather: Do I have good enough reason to pursue such a career myself?

---

[45] See "Reasons for Action: How to Reconcile the Autonomy and Heteronomy of Reason" (unpublished manuscript).

[46] This is my answer to Miller's *normative* defense of the priority of each person's own point of view. "Just as," Miller writes, "parents have a prerogative, within limits, to try to pass on their way of life to their children, I can reasonably reject a rule that requires me

It is easy to identify many considerations—both particular and general[47]—that seem relevant to answering this question. The hard part is to determine their relative significance.[48] For every consideration, there is a counter-consideration; for every argument there is a counterargument. To take just one example, it seems to me that a human life is better if it includes projects whose value does not itself depend on the fact that they satisfy human needs; human lives have a special value, it seems, because human beings can and do commit themselves to relatively unhelpful projects. The fact that we have reason to preserve human life thus seems to depend, at least in part, on the fact that human beings have the capacity to pursue projects that are not directly responsive to their needs; this capacity is an important aspect of human freedom. Even if I am right about this, however, it hardly follows that anyone ought to pursue relatively unhelpful projects *now*, or that *I* ought to do so at any time in my life. After all, satisfying one's most basic needs is a necessary condition for the possibility of living a better life. In this sense, surely, relatively helpful projects have a priority over those that are relatively unhelpful.[49]

---

to end the continuing presence of my current personality in my own life even though my goals are worthwhile" ("Beneficence, Duty, and Distance," p. 362). My point, I hope it is clear, is that even if Miller is right, he has not provided a justification for pursuing "worthwhile" goals that can be of no help to anyone in desperate need. As Herman points out, "Moralty cannot be so demanding that we are unable to live our lives.... And yet it is absurd, or worse, to stand on one's privilege in the face of need" ("Morality and Everyday Life," p. 42).

[47] It is important to stress that when we consider "how the reasons add up," we may well be convinced that we are constrained by certain very general principles.

[48] For one thoughtful attempt to do just this, see Robert Adams, *Finite and Infinite Goods* (New York: Oxford, 1999). According to Adams, Gauguin's case "illustrates the point that even if one has, or has had thus far, an ongoing vocation to love and pursue a certain project, it can be idolatrous, and terribly so, to refuse to set the project aside to attend to a more urgent duty or need." In other cases, however, including the case in which one is dedicated to philosophical research, Adams is "inclined to think" that one "may well be right, and in some such cases would be right, to pursue one's vocation anyway, at some cost to other people.... In the case of philosophical research,...it seems to me more plausible to hold that the vocation one believes one has is itself a ground of the rightness of the actions that it requires" (pp. 315–16). To my mind, this tentative suggestion invites more questions than it answers.

[49] Wallace, we have seen, believes that if eudaimonistic considerations conflicted with moral considerations, then the former would trump the latter. (See note 44.) Even if this normative assumption is correct, however—even if there is a conception of a "mea-

More generally, if I really do have reason to help other human beings in need, this must be grounded in what we philosophers rather antiseptically call "the value of humanity." Yet there are other ways to honor this value: Brahms did this, for example, when he wrote his Requiem, and so do those who sing the Requiem—and those who listen. The profusion of reasons for and against relatively unhelpful careers reflects, in large part, the profusion of ways in which it is possible to express love and respect and solidarity and compassion and joyful appreciation of the human kind of being: among these ways are not only all kinds of art, but also sports, and philosophy, and even good conversation. I have already conceded that these are all valuable human activities, and that their value is of different kinds, none of which can be reduced to the value of helping to save human beings from acute suffering and death. This in itself, I noted, makes it difficult to compare them. Yet insofar as they are all, in their way, responses to the value of humanity, the difficulty has another source: as far as I can tell, we do not really understand this very important value. Clearly, I do not really understand it myself.

The conclusion of my investigation thus has two parts: I do not believe that there is any deep justification for relatively unhelpful (though valuable) projects; I am uncertain about how to sort through the many competing (more or less general) considerations that are not so deep. Given how much is at stake, I surely have good reason to gain a clearer understanding of my options; I have good reason not to allow myself to be so muddled about the appropriate way to respond to the value of humanity. But, again, I seem to have good reason to do many other things, too. How much time and energy is it reasonable to devote to finding a principled way of sorting everything out?

The philosophical arguments in defense of my chosen career purport, in effect, to establish that I am justified in putting an end to

---

ningful life" that is sufficiently determinate to represent a constraint on what we have reason to do—it does not seem to provide an argument in support of minimally helpful careers. This is not only because, as Wallace himself is eager to point out, our interest in living a meaningful life seems to commit us to living a moral life, but—more basically—because just as I could be happy pursuing a more helpful career, so too, such a career could be part of a valuable, worthwhile, meaningful life. Indeed, this is why I could be happy under these circumstances. There are, of course, many interesting discussions of what is involved in living a meaningful life. In addition to Wallace's article, see Wolf's sustained exploration, in "Meaningful Lives in a Meaningless World," *Quaestiones Infinitae*, XIX (1997): 1–22 (publication of the Department of Philosophy, Utrecht University), "Happiness and Meaning: Two Aspects of the Good Life," *Social Philosophy and Policy*, XIV (Winter 1997): 207–25, and "Meaning and Morality," *Proceedings of the Aristotelian Society*, XCVII (1997): 299–315.

this inquiry and continuing along my inertial path. But these arguments do not take the inquiry seriously enough. As I have aimed to show by remaining in the first person, when one is trying to figure out what one has reason to do, one cannot bring one's deliberations to a close by simply reminding oneself of the role that commitments play in all reasoning, or of the fact that moral values are not the only kind, or the fact that Reason has its limits, and that once we have reached them, we must rely on extra-rational psychic resources. If, at this point in my own life, I have closed the career question, this is not because I am convinced that my choice is justified, nor that further inquiry would be a waste of time. It is simply that, for now anyway, I have run out of steam.

Necessarily, finite rational agents run out of steam at some point. And finite rational agency is the only possible kind. If this does not mean that we are justified in "just picking" any valuable project that strikes our fancy, what *does* it mean? Perhaps nothing, in the end. But I want to suggest—and this is, in effect, a further conclusion—that it may point to a necessary condition for the possibility of rational agency that is overlooked by attempts to turn rational agency into a source of standard moral obligations, or prerogatives, or both. More specifically, if still rather vaguely, I think it may point to an aspect of treating rational agents with respect that is obscured by the familiar focus on what we can rationally endorse.

"My apologies," writes the poet Wislawa Szymborska, "for the minuet record, to those calling out from the abyss. My apologies to those in train stations for sleeping soundly at five in the morning.... My apologies to large questions for small answers.... Truth, do not pay too much attention. Solemnity, be magnanimous toward me. Bear with me, O mystery of being, for pulling threads from your veil. Soul, don't blame me that I've got you so seldom."[50] Unlike the poets, we philosophers do not go in for addressing Truth. We would never think of doing anything so outrageous as to seek favors from the mystery of being! Yet the attitude expressed in this poem seems the right one: we need to be forbearing and magnanimous in response to those limitations of our rational agency which offer us neither justification nor excuse.

"I know," Szymborska continues, "that as long as I live, nothing can excuse me since I am my own obstacle."[51] Necessarily, a finite rational agent is its own obstacle—and, again, *finite*-rational agency is the only possible kind. Faced with an overwhelming profusion of values and valuable things, a rational agent must turn her back on most of them. This means that she must turn her back on the possibility that she has turned her back on the wrong ones. And if she has turned her back on the wrong values, she will almost surely have turned her back on human beings who deserve better.

---

[50] Szymborska, "Under a Certain Little Star," in *Miracle Fair,* Joanni Trzeciak, trans. (New York: Norton, 2001), on p. 114.

[51] "Under a Certain Little Star," pp. 114–15.

Most rational agents face this possibility without ambivalence—at least most of the time. In an obvious sense, this is a sign of sanity and maturity: there is no reason to wring one's hands over a necessary feature of rational agency; indeed, there is no reason to bemoan the fact that we are limited in many ways that go beyond the requirements of rational agency and that affect our assessments of what we have reason to do. But, again, the necessary conditions for the possibility of our acting for reasons are not themselves reasons for action. So, again, even if, given the limitations of agency, it is perfectly natural for us to feel pretty comfortable with the evaluative assumptions that get us through the day, it is always possible that these assumptions prevent us from seeing what we really have reason to do. This possibility—this essential fallibility of reason—warrants a forbearing and magnanimous attitude toward our choices.[52]

If a person is magnanimous, then she is open to the possibility that someone's actions may be justified, or at least excused, even if the appearances suggest otherwise. If a person is forbearing, then she refrains from reacting negatively when she thinks that someone has an excuse for behaving in a way that cannot be justified. But magnanimity and forbearance have a role to play even when someone's behavior can be neither justified nor excused. Under such circumstances, they take the form of a sympathetic response akin to forgiveness.[53]

It is often difficult to bring oneself to forgive someone who has behaved in a way that can be neither justified nor excused. But if the limitations of rational agency justify a forgiving attitude toward many evaluative errors, then one's magnanimity and forbearance should be directed, not only toward others, but also toward *oneself*. And this poses a very different sort of challenge: rather than having to work hard to overcome one's negative attitudes and feelings toward the behavior one would like to forgive, one must work hard to preserve the distinction between forgiving the behavior and judging that it can really be justified or excused, after all. To lose touch with this distinction is to fall into a trap: "[bad faith] stands forth in the firm resolution *not to demand too much,* to count itself satisfied when it is barely persuaded, to force itself in decisions to adhere to uncertain truths."[54]

---

[52] In case the reader needs an example to get the point, it will, I hope, suffice to remind her that rational agents who fly planes into tall buildings full of other (mortal) rational agents do not usually suffer from ambivalence. They pursue their projects with the firmest conviction that reason is on their side. If, moreover, it is easy to be mistaken about what is really worth doing, it is even easier to be confused about the *relative* merit of various valuable endeavors.

[53] One obvious difference is that forgiveness is an attitude that can only be taken by the one who has been wronged by the person being forgiven.

It is tempting to treat the fallibility of reason as a justification for paying very little attention to other human beings in need. But this would be to make the very mistake I have been arguing against in this paper. It would be to demand too little under the guise of not demanding too much. If, as I have argued, we cannot justify our choices by pointing out that there is no ur-value in terms of which all good things can be ranked, or by noting that decisions cannot possibly be made "from nowhere," this is because we cannot justify or excuse our choices by appealing to any limitation intrinsic to reasons and reasoning. If such limitations justified our actions, then *all* of our actions would be justified. If they excused our actions, then we would not be accountable for anything we do.

Whether we have an excuse for failing to appreciate that a given project cannot be justified depends on what, in particular, blinds us to the relevant considerations. But even if our ignorance, and our failures of imagination, can sometimes be excused, this is of no help to us as decision-makers. In deciding which projects to pursue, we want to know which projects can be justified. Our efforts to weigh the considerations for and against should inspire sympathy precisely because the task we face is so very demanding.

It is in the very nature of rational agency that it is inadequate to the task it sets itself: our reason cannot sort through all of the relevant reasons there are. It is thus in the very nature of rational agency that it needs to be supplemented with magnanimity and forbearance. Since the need is intrinsic to rational agency, the supplements are not merely supplemental. It is a contingent, though in large measure, remediable, fact that human beings are daily dying of malnutrition and disease.[55] But it is not a contingent fact that practical reasoners must ultimately "go with their gut," or "follow their heart," or just give in

---

[54] Sartre, *Being and Nothingness,* Hazel Barnes, trans. (New York: Washington Square, 1956), p. 113.

[55] In particular, it is a contingent fact that wealthy nations have done so little to address such needs at home and abroad. I hope it is obvious that I have offered no argument against the conviction that "where a community values something—art, or

ALIPORE CAMPUS

to inertia. Rational agents must eventually stop reasoning. So not paying as much attention as one might, not being as solemn, not having as much soul are all among the necessary hazards of rational agency. There is always a danger, too, of being too solemn about what is really not such a big deal, paying too much attention to what is really not so important; investing too much soul in the wrong things. We need to find the right way of accommodating this vulnerability without falling prey to the false comfort of self-justification and self-exculpation. We need to learn to settle for small answers to large questions without thinking that our answers make the questions irrelevant.

SARAH BUSS

University of Iowa

---

scientific innovation, or even sport—it can create conditions hospitable to creativity, and put in place both preventive and compensatory arrangements for those who may be negatively affected by these agents' moral negligence" (Herman, "Morality and Everyday Life," p. 40).

ERROR THROUGH MISIDENTIFICATION: SOME VARIETIES*

Memory provides normal human subjects with an ability of effortless, unreflective, mostly reliable judgment about the past. Judgments made in this way, like "I was in that remote part of Scotland five years ago," are not infallible, to be sure. Memory can play tricks. But it is plausible to suppose that memories are at least not liable to error through *misidentification*[1]—I can be wrong about being in that remote part of Scotland five years ago, but not in the way I can be wrong about "Aunt Miriam is wearing a red hat today," when I mistakenly take an approaching red-hatted woman to be my Aunt Miriam.

Suppose, however, that "quasi-memories"[2] are accepted as a coherent category, embracing genuine memories as a potentially proper subset, and that subjects could in principle be caused to have memory impressions deriving from someone else's bodily past. Such impressions might be subjectively indistinguishable from memories proper.

---

*I would like to thank Akeel Bilgrami, Carol Rovane, Crispin Wright, and an anonymous referee for helpful comments on previous versions of this paper.

[1] The label "(immunity to) error through misidentification" was first introduced in the philosophical literature by Sydney Shoemaker, "Self-reference and Self-awareness," this JOURNAL, LXV, 19 (October 3, 1968): 555–67, reprinted in Quassim Cassam, ed., *Self-Knowledge* (New York: Oxford, 1994), pp. 19–48 (page reference to the latter) in connection with Wittgenstein's distinction, between uses of "I" as subject and as object—*The Blue Book* (Oxford: Blackwell, 1958), pp. 66–67.

[2] The notion of quasi-memory is characterized by Gareth Evans as follows: "A subject *q*-remembers an event *e* if and only if (i) he has an apparent memory of such an event, and (ii) that apparent memory in fact embodies information deriving from the perception of that event by a person who is not necessarily himself"—*The Varieties of Reference* (New York: Oxford, 1982), pp. 247–48. It is assumed that one could have memory impressions deriving from someone else's past if one had the relevant brain cells transplanted into one's brain. Since quasi-memories and ordinary ones are phenomenologically indistinguishable, their phenomenological content is best expressed by locutions like "I seem to remember being in Scotland five years ago," or as "I have an (apparent) memory *as of* being in Scotland five years ago." It is then a further issue—an issue which I will not discuss in this paper—whether quasi-memories and ordinary ones should be taken to be intrinsically the same kind of mental state whose difference resides merely in their different causal origins, or whether—in a "disjunctivist" spirit—they should be taken, because of their different causal origin, as distinct and mutually exclusive mental states. For the purposes of this paper, we can be neutral on this issue. For once it is allowed that memories and quasi-memories are subjectively indistinguishable, it should also be allowed that the judgments based on either of them will have the same structure of *rational grounds*. There will be more about the notion of grounds in the sequel. But to help fix ideas, I take it that both in cases of memory-based self-ascriptions and in cases of quasi-memory-based ones, subjects requested to justify them would offer the following: "I seem to remember being *F*. Therefore, I was *F*."

0022-362X/06/0308/403–25

And according to Sydney Shoemaker,[3] their mere possibility suffices to show that ordinary memory-based self-ascriptions are *in principle* liable to error through misidentification. It is merely a contingency that they are not affected by it.

In contrast, according to Gareth Evans,[4] the possibility of quasi-memories, even if granted, shows nothing about ordinary memory, and judgments of the kind indicated remain *immune* to error through misidentification not merely in fact but *tout court*. I shall enlarge on Evans's grounds for this view shortly—here it will suffice to say that Evans takes it that misidentification, correctly so termed, should involve a false identity belief—in the example above, that that woman is Aunt Miriam. But in cases of judgment based on apparent memory, the identity of the subject seems to be built in from the start: perhaps I was not in that remote part of Scotland five years ago, but the false memory impression already identifies *me* as its subject. By contrast, in the Aunt Miriam case, there is a *correct* visual impression—that the woman in question is wearing a red hat—and it is her that I misidentify.

The immunity to error through misidentification of (some) self-ascriptions has been held to have very far-reaching metaphysical or semantic implications. Surely the self would have to be a very special kind of object to be unmistakeable in this way! Famously, Wittgenstein reacted in the opposite direction, taking the phenomenon to show that, in relevant cases, "I" is not used as a device of reference at all but is better assimilated to the merely grammatical subject of reports of the weather. G.E.M. Anscombe radicalized this view to apply to all self-ascriptions. In a similar vein, Shoemaker's and Evans's respective positions have been taken as the starting point of defenses of radically contrasting conceptions of persons and personal identity: to wit, the neo-Lockean view espoused by Derek Parfit and the "animalist" conception defended by John McDowell.[5]

---

[3] Shoemaker, "Persons and Their Pasts," in his *Identity, Cause, and Mind* (New York: Cambridge, 1970), pp. 19–48, here see pp. 23–24, 46.

[4] Evans, pp. 235–48, here see pp. 241–45.

[5] See Parfit, *Reasons and Persons* (New York: Oxford, 1984), pp. 220, 222; and McDowell, "Reductionism and the First Person," in Jonathan Dancy, ed., *Reading Parfit* (Malden, MA: Blackwell, 1997), pp. 230–50. According to Parfit, memories, as such, have an identity-neutral content and can give one information about the life of a specific human being only through the identification of oneself with the person whose life is responsible for the memories one is having. Such an identification might be wrong but, more importantly, it could be made by *different* future selves. This would be the case if, with an appropriate brain transplant, one's memories could survive in two (or more) different brains, belonging to two (or more) different human beings. Psychological continuity would then be preserved at the expense of personal identity.

What we face in this area is a vivid example of what Christopher Peacocke has recently called the "Integration Challenge," that is, the challenge of integrating metaphysics with epistemology. And surely both the Cartesian and the Wittgensteinian position provide us with examples of bad responses to the challenge. A prima facie peculiar aspect of the epistemology of self-ascriptions—the seeming (im)possibility of going wrong in identifying their subject (at least in some cases)—provokes a jump to "an exotic domain of the transcendent" or an implausible semantics, where what is needed instead is, in Peacocke's phrase, "revising and deepening one's epistemology."[6] The considerations to follow are offered in the spirit of that suggestion. Even if we are confident that immunity to error through misidentification is compatible both with the view that selves are identical to living human beings and that "I" is a genuinely referring expression,[7] the fact is that there has so far been little agreement either about the correct characterization of the phenomenon, or its extent.[8] And important aspects

---

The driving thought behind McDowell's position is that memories have immediate and irreducible first-person content. According to him, this is so because memories have the specific, biologically-selected function to retain, *for each specific human being*, information about his own past, and make it immediately available to him in the first-person mode of thought. Memories, therefore, are situated in the life of a specific human being and this latter fact, contrary to what Locke and Parfit seem to maintain, cannot be abstracted away in giving an account of persons and their identity. For an illuminating discussion of Parfit's and McDowell's positions, see Carol Rovane, "Personal Identity: Ethical Not Metaphysical," in Graham F. MacDonald and Cynthia MacDonald, eds., *McDowell and His Critics* (Malden, MA: Blackwell, 2006), forthcoming.

[6] Peacocke, *Being Known* (New York: Oxford, 1999), p. 263. The task of locating the exact source of such a challenge in the case of reflections on the self—whether in what Peacocke calls the phenomenon of "representational independence" or in something else—falls outside the scope of this paper.

[7] Evans, McDowell, Peacocke, and Crispin Wright in his "Self-knowledge: The Wittgensteinian Legacy," in Wright, Barry Smith, and Cynthia McDonald, eds., *Knowing Our Own Minds* (New York: Oxford, 1998), pp. 13–45, just to mention the most prominent figures, have all made these points. The arguments are usually that immunity to error through misidentification is common to both some uses of "I" and, for example, to perceptual demonstrative judgments, for which there is no temptation to think they may not refer; and that immunity is a function of the absence of any identification component in the rational grounds for a given first person judgment and not of the fact that either there is no self to be identified in the first place, or that the identification is secured by the special kind of entity presented to one for recognition.

[8] With respect to the former issue, we will see in the following that matters are far from being settled. With respect to the latter, nowadays nobody would agree with Wittgenstein's position that all and only psychological self-ascriptions are immune to error though misidentification (but for a different interpretation of Wittgenstein's position, see Brian Garrett, "Wittgenstein and the First Person," *Australasian Journal of Philosophy*, LXXIII (1995): 347–55). Since Shoemaker and Evans, it has become commonplace to hold that also nonpsychological self-ascriptions, such as those based on memory, perception, and somatic proprioception, as well as perception-based demon-

of the epistemology of the first-person judgments involved remain unclear. It is with the bearing of the Shoemaker-Evans dispute on these matters that I will be primarily concerned in this paper.

A useful discussion of some of the issues is James Pryor's.[9] His conclusion is that a proper understanding of the Shoemaker-Evans dispute should lead us to acknowledge two quite different kinds of error through misidentification—what he labels "*de re*" and "*which-object*" misidentification—involving respectively quite different structures of grounds and different kinds of mistaken belief. Below I shall argue that Pryor's distinction is misguided in detail. However, I accept—and shall argue—that there is a (different kind of) distinction to draw, and that drawing it will allow us to reconcile the positions of Shoemaker and Evans. Let me start by outlining the Shoemaker-Evans dispute in more detail.

THE SHOEMAKER-EVANS DISPUTE AND PRYOR'S RESOLUTION

Surely there is something right about Shoemaker's view. Once one accepts the notion of quasi-memory, it seems that I might have apparent memories as of, say, being in a particular remote part of Scotland five years ago, which in fact causally derive from someone else's past. And in such an event, according to Shoemaker, my judgment, "I was in that remote part of Scotland five years ago," would be affected by an error of misidentification. There would be *someone* who was in that remote part of Scotland five years ago, and I would be right about that, but wrong about whom. Yet Evans's counter seems powerful too. It is that even in those abnormal conditions, there would be no "identity-neutral" memory content, such as "I seem to remember *someone* being in that remote part of Scotland five years ago," which could lead to the relevant self-ascription only via a collateral identity belief such as "I = the person whose past I seem to remember now." Indeed, it is because the very content of my apparent memories is that

---

strative judgments are immune to error through misidentification. Moreover, since Evans (see pp. 219–20), it has been widely held that *inferential* psychological self-ascriptions are liable to error through misidentification. Finally, John Campbell by appealing to the phenomenon of "thought insertion", whereby schizophrenic patients claim that they have a given thought which is not theirs, has gone so far as to argue that also *noninferential* ones can be liable to such an error—see "Schizophrenia: The Space of Reasons and Thinking as a Motor Process," *The Monist*, LXXXII (1999): 609–25. (For a discussion critical of Campbell's position, see my "Thought Insertion and Immunity to Error through Misidentification," and "What There Really Is to Our Notion of the Ownership of a Thought," *Philosophy, Psychology and Psychiatry*, IX, 1 (2002): 27–34, 41–46; and Campbell's reply "The Ownership of Thought," *Philosophy, Psychology and Psychiatry*, IX, 1 (2002): 35–39.)

[9] Pryor, "Immunity to Error through Misidentification," *Philosophical Topics*, XXVI, 1 (1998): 271–304.

I was in that place five years ago that I do not require collateral information bearing on whether I am identical to the person whose past is responsible for them, in order to be rationally justified in holding that I was. However, if my self-ascription is not arrived at and supported by my holding any suitable identity belief, then, according to Evans, there is no scope for the former to be affected by error through misidentification.[10] No doubt such a self-ascription would be mistaken in the kind of case envisaged—I *was not* in that remote part of Scotland, after all. Still, it would not be mistaken because I have *misidentified* the person who instantiates the property of having been in that place five years ago, but, rather, because I am suffering from the *illusion* of having myself been in the place in question five years ago. I am not misidentifying anybody, but merely attributing to myself a *property* that I do not have.[11]

It would perhaps be an overstatement to call this opposition a paradox, but it is undeniable that both theorists seem to have a point. There is indeed a strong intuition that, as Evans holds, memory impressions, no matter what their causal origin might be, have irreducible first-personal content and that the self-ascriptions based on them, being free of reliance on any collateral identification, are therefore immune to error through misidentification. Still, it is equally intuitive to hold, with Shoemaker, that in the abnormal conditions envisaged, there *would* be an error of misidentification. After all, there would be someone who has the property of having been in that remote part of Scotland five years ago—it is just that the person in question is not whom I take him/her to be. So it would be a welcome result if *both* these intuitions could somehow be respected. Indeed, I submit, it is exactly because Shoemaker's position has been taken to imply the denial of the fact that memory impressions, no matter what their causes might be, have irreducible first-person content that it has been contested.[12] Yet the point remains that Evans's position too involves denying a prima facie powerful intuition: that in the abnormal conditions envisaged there would be an error not of predication—after all I rightly judge that someone was in that remote part of Scotland five years ago—but, rather, of identification of the person who instantiates that property.

---

[10] See Evans, pp. 235–48; McDowell, pp. 239–41; Wright, p. 20.

[11] Another reason why Evans has often been taken to prevail in the dispute is that he construed error through misidentification as a case in which although one's final judgment may be mistaken, one still retains knowledge that *some* (*other*) person was (or appeared to be) *F*. However, in deviant causal conditions, a subject who is storing memory information from someone else's past cannot read off, from the content of his apparent memories, that *someone else* was *F*, precisely because their content is as if *he* was *F*. As Evans put it, in discussing the parallel case of self-ascriptions based on somatic proprioception: "If the subject does not know that *he* has his legs bent (say) on this basis (because he is in the situation described, [that is, he is receiving proprioceptive information from someone else's body], then he does not know *anything* on this basis. (To judge that *someone* has his legs bent would be a wild shot in the dark)" (p. 221, see also pp. 244–45). Notice, however, that even so, the subject would still be a reliable indicator of someone else's past. Hence, if he was given the additional information that he is storing information from someone else's past he would, by means of his apparent memories, gain knowledge of someone else's past.

Now, Evans allowed for the possibility of informing the subject of the fact that he could store information deriving from someone else's past. But, in that case, he thought that the subject would also be "in the position of knowing that the information was not being received in the *normal way*" (p. 27n [emphasis mine]). So he agreed that the relevant identification-free judgments rest on an assumption, namely, the assumption that the information underpinning them is stored in the usual ways. But he denied that they rested on an *identity* assumption (p. 186). That is to say, he denied that they rested on an assumption such as "I = the person whose past I seem to remember now." There will be more about such identity assumptions and their precise role in licensing the relevant judgments in the sequel.

Pryor's suggestion is accordingly welcome. It is that distinguishing between two different species of error through misidentification—his respective categories of *de re* misidentification and *which-object* misidentification[13]—lets us settle the conflict. For although it is true that memory-based self-ascriptions cannot be afflicted by the former kind of error, because they are based on irreducibly first-personal contents and are, therefore, identification-free, just as Evans says, they can nevertheless be afflicted by the latter, since, as we shall presently see, this variety of error through misidentification supposedly does not involve a belief in any (mistaken) identity judgment.[14] Pryor's own announced intention is to vindicate Shoemaker's position against Evans.[15] But in effect what he suggests is a way of doing justice to both their respective intuitions and a possible *dissolution* of their disagreement. For if we accept his proposed distinction we can say that while Evans was right on the former—allegedly less interesting[16]—understanding of error through misidentification, Shoemaker too was right, on the latter—allegedly more interesting—understanding of it.

---

[12] See, in particular, Evans, pp. 240–48; and McDowell. It should be noticed, however, that while Shoemaker himself held the view that memory-based self-ascriptions are grounded in *first-personal* memory impressions (Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca: Cornell, 1963), pp. 33–34) and yet, in the abnormal conditions envisaged, may be liable to error through misidentification (Shoemaker, "Introspection and the Self," in his *The First Person Perspective and Other Essays* (New York: Cambridge, 1986), pp. 3–24, see p. 15), it is in fact Parfit (pp. 220, 222), who has contributed to spreading the view that memory-based self-ascriptions are always based on *identity-neutral* memory impressions and arrived at by holding in place the relevant identity beliefs.

[13] See Pryor, pp. 274–76, 282, 284.

[14] Friends of disjunctivism about memory should read Pryor's claim as being about the grounds for I-judgments based on (apparent) memories. Accordingly, while I-judgments based on (apparent) memories would not be grounded in identificatory beliefs, they would nevertheless be grounded in beliefs such as "I seem to remember being in Scotland five years ago," that remain liable to which-misidentification.

[15] Pryor, p. 272: "I will show how we can put these distinctions to use in assessing a debate between Shoemaker and Evans about whether first-person memory-based beliefs are immune to error through misidentification. I will argue that, with respect to the most interesting sort of immunity, Shoemaker is right: our first-person memory-based beliefs do *not* have that sort of immunity."

PRYOR'S TWO VARIETIES

Let us turn, then, to the detail of Pryor's proposal. Consider the Aunt Miriam example. Suppose I am at a car park and see a woman with a red hat approaching. I mistakenly take that woman to be my Aunt Miriam. Hence, I form the judgment "Aunt Miriam is wearing a red hat." My judgment can be seen as rationally grounded on a pair of beliefs, namely, the belief "That woman is wearing a red hat" and the belief "That woman = Aunt Miriam." Call the former the *predication component* and the latter the *identification component.* Error through misidentification is naturally viewed as arising when, in a case of this general structure, the identification component is mistaken.[17]

This, I submit, is the paradigm case. The following three features are salient in it:

(1) Paradigmatic error through misidentification proper does not involve *mispredication.* It involves neither a mistaken impression of *what sort* of object one is presented with, nor a mistaken attribution of a given *property* to that (successfully targeted) object. Rather, it involves the mistaken identification of that object with another—the object of one's eventual judgment.

(2) In consequence, error through misidentification can arise only if one's grounds allow one to make *an independent singular judgment* (in the example, "That woman is wearing a red hat"). For since error through misidentification, properly so regarded, consists in the mistaken identification of the object, which instantiates a given property *F*, the grounds at one's disposal in making the judgment must be such as first to license a singular judgment about that object.

(3) Finally, when paradigmatic error through misidentification occurs, the *very same grounds* that support the singular judgment will, if the eventual judgment is defeated, *survive as grounds for a corresponding existential generalization.* In our example: my visual grounds for "Aunt Miriam is wearing a red hat" are such that, although that judgment is defeated—since, as it turns out, the person I am seeing is not Aunt

---

[16] See the previous footnote.

[17] Evans, chapters 6, 7; and Peacocke, *Sense and Content* (New York: Oxford, 1983), pp. 143–44.

> Miriam—they remain as grounds for "Someone (else) is wearing a red hat."[18]

Mindful of these three considerations, let us move directly to review Pryor's *second* kind of error through misidentification—"*which-misidentification.*" (Pryor's first kind of error through misidentification—the *de re* one—is relevantly close to the paradigm case just reviewed.) Pryor offers the following example to illustrate the general idea:

> I smell a skunky odor, and see several animals rummaging around in my garden. None of them has the characteristic white stripes of a skunk, but I believe that some skunks lack these stripes. Approaching closer and sniffing, I form the belief, of the smallest of these animals, that it is a skunk in my garden. This belief is mistaken. There is a skunk in my garden, but it is not the small animal I see.[19]

The epistemological structure of the case, Pryor suggests, is as follows:[20]

(i) A subject has some grounds *G* that offer him knowledge of the existential generalization that something is *F*.
(ii) Partly on the basis of *G*, the subject is also justified, or takes himself to be justified, in believing of some further object *a* that *it* is *F*.
(iii) But in fact *a* is not *F*. Some distinct object *b* is *F*, and it is because the grounds *G* "derive" in the right way from this fact *about b* that they offer the subject knowledge that something is *F*.[21]

---

[18] Note that it is crucial to this point that the grounds considered be exactly those that support the independent singular judgment. For, if we allowed for grounds other than the ones that support the independent singular judgment to license the relevant existential generalizations, it would then be possible to turn judgments into a case of error through misidentification *ad libitum*. To illustrate: suppose that I have *prior* testimonial grounds for the belief "Someone (in the car park) is wearing a red hat." I look and see Aunt Miriam seemingly wearing a red hat. I *then* judge "Aunt Miriam is wearing a red hat." However, as it turns out, Aunt Miriam is not wearing a red hat but a dark orange one. My visual grounds for the singular judgment "Aunt Miriam is wearing a red hat" would be defeated, yet my *antecedent* testimonial grounds would survive as grounds for "Someone (else in the car park) is wearing a red hat." Obviously, however, my retained entitlement to the existential generalization would not turn my singular judgment "Aunt Miriam is wearing a red hat" into a case of error through misidentification properly so regarded.

[19] Pryor, p. 281.

[20] In fact, Pryor's which-misidentification is partly reminiscent of Wright's characterization of error through misidentification, which is as follows:

> A judgment of the form "a is F" is liable to error through misidentification iff its grounds are such that, if defeated, they would survive as grounds for the corresponding existential generalization (Wright, p. 19).

However, while Wright requires the very grounds for the singular judgment to be such as to survive as grounds for a corresponding existential generalization, once defeated, Pryor does not.

[21] Pryor, p. 282 (with minor changes in the notation). This characterization is slightly revised later on in the paper (p. 284) in order to account for different kinds of defeating

In Pryor's view, then, *which-misidentification* does not require that one's grounds for the eventual judgment include an identification component at all. The key difference between it and the paradigm above is that while, in the latter case, the grounds I have for my judgment justify me in believing *of* some particular object *b* that it is *F*, and I mistakenly take *b* to be identical with *a*, in a case of *which-misidentification* my grounds do *not* justify me in believing *of a particular object b* that it is *F* but only that *something* is *F*. As Pryor puts it: "In cases of *which*-misidentification, I go wrong not in *re-identifying* the thing I know to be *F* as some other thing; rather I go wrong in figuring out *which thing* is *F* in the first place."[22]

It is then this feature of which-misidentification that, as Pryor intends, allows us to respect, on the one hand, Evans's intuition that memory-based judgments are identification-free, while acknowledging, on the other hand, Shoemaker's intuition that these judgments can be affected by error through misidentification when one is in fact storing memory information deriving from someone else's past. Simply: such judgments, in Pryor's view, are immune to *de re* misidentification because they do not rest on any identification component, but are open, in principle, to which-misidentification.

Strikingly, an immediate problem with Pryor's proposal is that even if he were right in inviting us to distinguish between these two kinds of error through misidentification, his template—and clause (i) in particular—will allow which-misidentification to apply to memory-based judgments only if their grounds are identity-*neutral*. My apparent memory would have to be conceived as first giving grounds for the existential claim "Someone was in that remote part of Scotland five years ago" which would then have to be somehow connected to the eventual judgment about myself. So the proposal in effect merely invites a restatement of Evans's original objection—that it simply distorts the epistemology involved to suppose that memory provides such identity-neutral grounds.

Be that as it may, a deeper problem with Pryor's proposal emerges. Reflect that, on his view, the skunk example is said to be a case of error through misidentification because it involves identification of one object rather than another as the instance that verifies the

---

evidence and for the possibility of the falsity of the existential generalization. For present purposes, the first, simpler characterization of which-misidentification will suffice.

[22] Pryor, p. 283.

existential quantification for which the thinker has independent grounds.[23] But how exactly, in the circumstances of the example, is *that* identification arrived at? On reflection it seems evident that the rational grounds underlying the identification may as well be conceived as follows. I smell a skunky odor and this gives me grounds for:

(1) The animal (in my garden) which is actually responsible for this odor I can smell is a skunk.

I go closer to one of the animals and sniff, while I can still smell the skunky odor. I then judge "This animal is a skunk" *because* I believe:

(2) This animal (I can now see) = the animal (in my garden) which is actually responsible for this odor I can smell.

So the resulting judgment—that this animal is a skunk—now has a similar structure of grounds to the paradigm case, and since the animal that is visually presented to me is not the one that is actually responsible for the odor I can smell (nor is it a skunk), my judgment is false precisely because the identification component, (2), is mistaken.

Once the case is thus represented, it exhibits all the characteristic features of error through misidentification in the paradigm case. It is a case of misidentification of the animal which is causally responsible for the skunky odor I can smell. Moreover, the olfactory grounds at my disposal allow me to make a *singular* judgment and my visual grounds, in turn, allow me to identify one particular object as the one which my olfactory grounds represent as having the property of being a skunk. Finally, the defeat of (2)—the identification component—is compatible with the retention of the original (olfactory) grounds for the belief in the *singular* predication component (1), "The animal (in my garden) which is actually responsible for this odor I can smell is a skunk," from which the existential generalization follows.[24]

I therefore want to grant that Pryor's paradigm example of which-misidentification may indeed be construed as a case of error through misidentification. But, when it is, the claim that it illustrates a different *kind* of error through misidentification, which does not involve as part of one's grounds a belief in an identification component, is wrong. For, when it is properly reconstructed, the example offered by Pryor is merely one further illustration of the usual structure of grounds underpinning error through misidentification.

---

[23] I would like to thank an anonymous referee for making me see this point.

[24] The fact that Pryor's example involves *mixed* grounds shows that if one wished to read it as a case of error through misidentification properly so regarded then the following features will be salient. *All* grounds at a subject's disposal in forming the relevant judgment are such that they independently sustain a singular judgment and, were they to be defeated, they would remain as grounds for the *singular* predication component, *from which* the existential generalization would follow.

The source of Pryor's mistake lies, I suggest, in too narrow a conception of the range of *concepts* which may feature in an identification component. An identification component may involve either *de re* concepts—that is, concepts which are grounded in identifying knowledge of their unique instances, like the concepts "That person" and "Aunt Miriam" in the recognition-based example I started with. Or else, it may involve non-*de re* but still singular concepts (hereafter, just "singular concepts") that are not grounded in identifying knowledge of their unique instances, like "The animal (in my garden) which is actually responsible for this skunky odor I can smell." The difference intuitively is this. Our thought can be directed onto a *specific* object in two different ways. It can be directed onto it by taking it to be the unique object that satisfies a given property—in Pryor's example, the property of being the animal in my garden that is actually responsible for the skunky odor I can smell. Or else, it can be directed onto it in virtue of the fact that a subject has identifying knowledge of it—knowledge which, in typical cases, enables him to identify it *perceptually* among other objects.[25]

So it is true, as Pryor has it, that when I make an error through misidentification in the latter kind of case, I go wrong in *reidentifying* a, because I wrongly take b to be the single instance of the corresponding *de re* concept "a." Similarly, it is true, as Pryor suggests, that in the former case, I go wrong in *identifying which* object is the one that satisfies a given uniquely identifying property. For instance, in our example, I mistakenly take *this* animal I can see in front of me as the one which is actually responsible for the skunky odor I can smell. Still, there is no difference in the *structure* of the grounds underpinning my final judgment. In both cases they involve a belief in a false identification component. The difference resides, rather, in the character of the concepts configured in that component.

---

[25] It may well be that if one allows for de re *mathematical* concepts (see, for instance, Saul Kripke's unpublished but widely circulated Whitehead Lectures on de re mathematical knowledge), then the identifying knowledge that would enable one to have such concepts would not depend on any kind of perceptual, or perceptual-like identification of the relevant objects. The intuitive thought behind the idea that there may be both de re empirical concepts and de re mathematical ones is this: in both the empirical and in the mathematical case, for certain instances of x, the question "The number of Fs is x, but which number is x?" and the question "x is F (or the F), but which object/person is x?" would not make sense. For present purposes, however, it is enough if being able perceptually to identify a material object is a sufficient condition for (and the normal way of) having de re concepts of *empirical* objects.

Notice, however, that the mathematical case is relevant in order to motivate the view that genuine identification components can contain not only de re, but also singular concepts. For in mathematics this would be the usual case. For instance, "$7+5=12$" or "$\sqrt{9}=3$" would all be identity judgments involving a singular concept on the left hand side—"the number which is the sum of 7 and 5" or "the number which is the absolute value of the square root of 9"—and a de re concept on the right hand side—"12" or "3."

To sum up: error through misidentification, I am proposing, properly construed, *always* depends on the presence of a false identification component as part of the grounds for a given judgment. Yet, the *constituents* of such a component may vary, being either *de re* concepts, or merely singular ones. Pryor's error, I want to suggest, consists in his interpreting the possibly different *nature of the conceptual ingredients* featuring in one's grounds for the relevant judgments as a difference in the *structure of those grounds.*

Unfortunately, however, it is rapidly evident that even this emended version of Pryor's distinction no longer subserves the thought that Shoemaker was right to take the kind of error that might be induced by a mere quasi-memory as an error of misidentification—as an error resting on a belief in a false-identification component. For, once again, what would the false identification component be? Suppose I merely quasi-remember being in that remote part of Scotland five years ago. Then the structure of grounds that would let us assimilate the example to the skunk case would be somewhat as follows:

(1) The person whose past I seem to remember now was in that remote part of Scotland five years ago.
(2) I = The person whose past I seem to remember now.
(3) I was in that remote part of Scotland five years ago.

If my judgment of (3) was indeed grounded in this way, and mistaken because (2) was mistaken, then it would involve error through misidentification in the corrected sense latterly elicited from our criticism of Pryor's notion of which-misidentification. But it appears once again a decisive point in Evans's favor that this utterly misrepresents the way in which ordinary memory claims are normally grounded, in which no role is played by identification components on the model of (2). The issue was whether, without tampering with that ordinary structure of grounds, the mere possibility of quasi-memory would introduce scope for error through misidentification. Pryor's case for an affirmative answer does not survive. For we have seen first, that his notion of which-misidentification is spurious; second, that in any case it does not deliver the intended result without serious distortion of the grounds of the memory judgments typified by our leading example; and third, that a correct account of the intent of his principal example can be superimposed on memory judgments only, again, at the expense of serious distortion of their grounds.

It remains to be seen how something in the spirit of Pryor's proposal may yet be redeemed.

VARIETIES OF GROUNDS AND OF ERROR THROUGH MISIDENTIFICATION

The previous discussion has urged that error through misidentification arises when and only when a judgment is rationally based on a mistaken identification component.[26] Liability to error through misidentification is accordingly a squarely *epistemological* phenomenon: it is intrinsic to the character of the grounds a thinker possesses for a judgment, and belongs to its subject matter or semantics only in so far as they restrict its possible grounds to those of the appropriate kind. This claim is accepted by much discussion in this area;[27] in particular—it is accepted by both Pryor and me.[28] Although, as we have seen, he thinks that something properly described as misidentification may occur in cases where there is no identification component among the grounds, and I reject this. But the reader may at this stage be puzzled. For have I not just signaled an intention to make a proposal about the Evans-Shoemaker dispute broadly similar in spirit to Pryor's? And how is that to be done if not by contrasting two different kinds of error through misidentification? And how is such a contrast to be made if it is insisted that all cases of such error share a common structure of grounds?

Well, I do in fact want to accept a qualification of that latter point: while all cases of error through misidentification, properly so viewed, do indeed involve a mistaken identification component, it is also true—as I shall later illustrate—that a mistaken identification component can nevertheless feature in *differently structured* sets of grounds. But that point is not the basis of the distinction I want to draw. Rather, I want to focus on what it means to say that a judgment which is liable to error through misidentification is *rationally grounded* in certain component beliefs or assumptions. My central contention will be that there are two intuitively quite different ways in which one judgment, or assumption, can ground another and that this distinction gives rise in turn to two potentially different ways in which an identification component can feature in the grounds for a given judgment, and thereby to two different kinds of error through misidentification. It will be the distinction so drawn that I think is of service in resolving the Shoemaker-Evans dispute.

---

[26] See, for instance, Evans, chapters 6, 7; Peacocke, *Sense and Content*, pp. 143–44.

[27] Evans, McDowell, Peacocke, Pryor, and Wright.

[28] The point is not taken universally. In my "Error through Misidentification, the Split between Speaker's and Semantic Reference and the Real Guarantee," this JOURNAL, C, 8 (August 2003): 216–31, I discuss extensively those interpretations of error through misidentification which take it to be a *referential* mistake, such as Elisabeth Anscombe, "The First Person," in Simon Guttenplan, ed., *Mind and Language: Wolfson College Lectures 1974* (New York: Oxford, 1975), pp. 45–64; Andrea Christofidou, "First Person: The Demand for Identification-free Self-Reference," this JOURNAL, XCII, 4 (April 1995): 223–34; and Rovane, *The Bounds of Agency* (Princeton: University Press, 1998). On these latter views, when that error occurs, in using "*a*" one would in fact refer to the person visually presented to one, rather than to its semantic referent. So, in the Aunt Miriam example, when I am in the car park and I mistake a red-hatted woman approaching for Aunt Miriam, by saying "Aunt Miriam is wearing a red hat" I would be (speaker-)referring to the woman who is actually approaching and not to my aunt. Since they do not find this outcome plausible in the case of the use of "I," they conclude that self-ascriptions are (almost) always immune to error through misidentification. In the paper I put forward the view that such an interpretation might in fact be based on a conflation between (immunity to) error through misidentification, the (im)possibility of a split between speaker's and semantic reference and what I label the "real guarantee": namely, the fact that any competent use of "I" is such that the speaker knows which object is its semantic referent, namely, *oneself*.

Intuitively, we can distinguish at least three different kinds of roles items of information can play in providing the *grounds* of a given judgment *P*:

(*) The information may provide the premises for a *conscious inference* to *P*.

(I) The information, while not part of a conscious inference, may be what a particular (rational and appropriately conceptually equipped) subject would appeal to if he were requested to justify his judgment that *P*. In such a case, I shall say that the information enters into *the subject's own rational grounds* for the judgment.[29]

(II) The information, while neither entertained in a process of conscious inference, nor part of the subject's own rational grounds—part of what he would appeal to in order to justify his final judgment—may be such that, were it somehow to fall into question, a (rational and appropriately conceptually equipped) subject would be prepared to withdraw from the judgment that *P*. In such a case, I shall say that the information provides certain *background presuppositions* on which the subject's final judgment rests.

We can set cases of kind (*) aside. For they involve no important epistemological differences from those of kind (I), contrasting only in whether the premises are used by a subject to arrive at his final judgment or to justify it *ex post*. The important difference is between cases of kind (I) and (II)—between rational grounds and presuppositions. To illustrate it, consider the following example.

(I) I see someone in the distance and, immediately taking her to be my Aunt Miriam, judge "Aunt Miriam is wearing a red hat." If, then, I had to justify my judgment, I would appeal to the relevant identification and predication components respectively ("Look, that's Aunt Miriam—and see, she is wearing a red hat"). In such a case, my judgment would be grounded in those components, although it was not arrived at through a process of conscious inference involving them.

---

[29] A somewhat similar point is made by Peacocke, *Sense and Content*, pp. 143–44.

(II) I again form the recognition-based judgment "Aunt Miriam is wearing a red hat" immediately. But if I were now to learn that, as a matter of fact, Aunt Miriam has an absolutely identical twin sister, nobody has ever heard of, and that she is in the neighborhood, I would be prepared to withdraw from my judgment. So my judgment rested on the background presupposition that Aunt Miriam is immediately visually recognizable—a presupposition that is defeated by learning that she has an absolutely identical, local twin sister. Plausibly, however, in order to be justified in forming the original recognition-based judgment "Aunt Miriam is wearing a red hat," it is not required that I (be able to) enumerate all its implicit background presuppositions, nor find out whether they have been met. They are taken for granted. Still, if any of them were somehow subsequently called into doubt, that would defeat my justification for the judgment.[30]

My proposal is, simply, to apply this distinction to the case of error through misidentification. Depending on *how* a (mistaken) identification component actually features, a judgment can be affected by an error through misidentification that occurs either (I) in the subject's own grounds for it, or (II) in its background presuppositions.

Construing the role of the identification component as in (I) will contribute towards accounting for error through misidentification *from the perspective of the subject.* For, on such a construal, despite the fact that a subject may not have gone through any process of conscious inference, he himself would appeal to the—as it happens, mistaken—identification component if he were asked to justify his own judgment. This contrasts sharply with cases of kind (II). Nonetheless, construing the role of the identification component as in (II) can offer an important insight into the part that component may play in the epistemic structure underlying *certain* judgments—characteristically, judgments we can make about our bodies, our physical location and—crucially for the Shoemaker-Evans dispute—our past.[31] Let me briefly explain what I take this insight to be.

---

[30] I am relying on the fact that the distinction between a subject's grounds and a judgment's background presuppositions is evident, although, of course, its further clarification may require additional work. Notice, moreover, that such a distinction is ubiquitous. Take for instance the judgment "Here's my hand." A subject's own grounds for it will, presumably, be given by his perceptual experience. Its background presuppositions, in contrast, will be that there is an external world, that he is not now dreaming, that his sense organs are working properly, and so on. So, were them to fall into doubt, he would have to withdraw from his final judgment. Still, this does not mean that he should be able to enumerate them and make sure they have been met in advance of forming his judgment.

For a discussion of whether the background presuppositions need in turn to be warranted (as such and not by the actual subject who is making a perceptual judgment) in order for the final judgment to be warranted, or whether it is enough for them not to be substantially questioned or doubted, see Wright, "Facts and Certainty," *Proceedings of the British Academy,* LXXI (1985): 429–72 and his "(Anti)-sceptics, Simple and Subtle: Moore and McDowell," *Philosophy and Phenomenological Research,* LXV (2002): 330–48; and Pryor's "The Skeptic and the Dogmatist," *Noûs,* XXIV (2000): 517–49—here see pp. 537–38, respectively. So, in my view, although they do not explicitly distinguish between the subject's own grounds for a given judgment and the background presuppositions of it, the best way of interpreting their dispute is to read it as bearing on whether background presuppositions need (or need not) be warranted.

Consider again my judgment, "I was in that remote part of Scotland five years ago," when based on memory. In such a case my own rational grounds would consist in—and be exhausted by—my apparent memory of my being in that remote part of Scotland five years ago. Yet, that apparent memory can provide a warrant for "I *was* in that remote part of Scotland five years ago" only so long as the identity, "I am the person whose past I seem to remember now," is not called into doubt. Obviously, my final judgment need neither be inferred from, nor—from my own point of view—rationally grounded in that identity. Yet, it does rest on it as a presupposition. Were that identity to fall into doubt, I ought to withdraw from my final judgment (even though, to stress again, the fact that a subject should withdraw from a judgment, once apprised of the questionability of certain claims, does not suffice to make those claims into part of his own rational grounds for that judgment; he may never have considered them and may have no specific evidence in their favor).

To illustrate the two different roles identification components can play, let us contrast the judgment "I was in that remote part of Scotland five years ago," made on the basis of my memories, with the same judgment made on the basis of finding out a piece of information in a diary, after, for instance, having lost any recollection of a visit to Scotland. In the latter case, the self-ascription could have the following grounds:

(1) The person who wrote this diary was in that remote part of Scotland five years ago.
(2) I am the person who wrote this diary.
(3) I was in that remote part of Scotland five years ago.

---

[31] It should be kept in mind that in the literature on error and immunity to error through misidentification cases of self-ascriptions based on personal memories like "I was in that remote part of Scotland five years ago," on somatic proprioception, such as "My legs are bent," and on the perception of one's environment, such as "I am in front of my car" are usually treated on a par. I will not expound on the latter cases. Arguably, however, they will rest on the background presupposition "I = the person from whose body I am now receiving perceptual/proprioceptive information."

Here, the role of the identification component—(2)—is indeed that of rationally grounding the *self*-ascription of the property of having been in that remote part of Scotland five years ago, which would otherwise be rationally unjustified. If I did not believe (2), I would not be rationally justified in self-ascribing the property of having been in that remote part of Scotland. I would only be justified in believing that *someone*—the person who wrote the diary—was in that remote part of Scotland five years ago.

Contrast the case in which "I was in that remote part of Scotland five years ago" is once again based on my current memory impressions. Here, as remarked, the structure of the grounds is:

(1) I seem to remember being in that remote part of Scotland five years ago.
(2) I was in that remote part of Scotland five years ago.

And the crucial thing to notice, just as Evans insists, is that premise (1) already involves a *self*-ascription, since I seem to remember being *myself* in that remote part of Scotland five years ago. The role of the background presupposition

(0) I am the person whose past I seem to remember now.

is merely that of sustaining the passage from the self-ascription of the property of *seeming* to remember being in that remote part of Scotland five years ago to the self-ascription of the property of *actually having been* in that remote part of Scotland five years ago. So, the presence of identification components featuring merely as background presuppositions allows for the possibility of judgments—like "I was in that remote part of Scotland five years ago" when based on one's memories—which, despite being grounded in irreducibly first-personal contents, and having no identification component among their grounds at all from the subject's point of view, may in principle yet presuppose a false identification.

THE RESOLUTION OF THE DISPUTE I: TWO VARIETIES

It remains to review the disagreement between Shoemaker and Evans more explicitly in light of this proposal. The question we must address is whether it is possible to rescue Shoemaker's position, once it is acknowledged with Evans—as I do—that (apparent) memories have irreducible first-person content and that judgments based on them are not rationally grounded on any identification component. The suggestion for which I have been preparing allows of two versions. On the one hand, one might propose that there are, after all, *two varieties* of error through misidentification—respectively, error through misidentification relative to one's grounds, and error through misidentification relative to background presuppositions—which can be defined as follows.

> Error through misidentification relative to the subject's own rational grounds: a judgment of the form "*a* is *F*" is affected by error through misidentification if and only if the subject's own rational grounds for that judgment contain a mistaken identification component.
>
> Error though misidentification relative to background presuppositions: a judgment of the form "*a* is *F*" is affected by error through misidentification if and only if a mistaken identification component features as part of its background presuppositions.

Two corresponding varieties of *immunity* to error through misidentification would then be characterized thus:

> Immunity to error through misidentification relative to the subject's own rational grounds: a judgment of the form "*a* is *F*" is immune to error through misidentification if and only if the subject's own rational grounds for that judgment do not contain any identification component.
>
> Immunity to error through misidentification relative to background presuppositions: a judgment of the form "*a* is *F*" is immune to error through misidentification if and only if no identification component features as part of its background presuppositions.

One may then argue that while memory-based judgments are in fact immune to error through misidentification *relative to the subject's own rational grounds*—since the latter are based on irreducibly first-personal contents and, therefore, do not contain any identification component—they remain liable to error through misidentification *relative to their background presuppositions*, in so far as the latter contain identification components that, in certain—no doubt remote—circumstances, might be called into question or substantially doubted. Evans would then be right on the former understanding of immunity to error through misidentification, while Shoemaker would be right on the latter.

THE RESOLUTION OF THE DISPUTE II: MERELY ONE VARIETY

That has a pleasing simplicity. One may worry, however, about a notion of error that does not require the *subject's* going wrong about anything—indeed, a notion of error that allows that a thinker may be mistaken through misidentification even though possibly lacking the conceptual resources to entertain the relevant belief. But just that is provided for by the characterized notion of error though misidentification relative to background presuppositions. One may therefore feel inclined to insist that it is best to countenance merely *one* variety of error through misidentification—that is, error relative to *the subject's* *own rational grounds* for his judgment—and that Shoemaker's position rests in effect on a confusion of presuppositions and grounds. On this view, there would be no scope for redeeming Shoemaker's view: if memory-based self-ascriptions are based on irreducibly first-personal contents, and therefore involve no relevant identification component in their rational grounds, then in the metaphysically abnormal, so to say, condition of quasi-memory one's judgment will be false alright, but will not involve any mistake of identification.

This, however, would be a hasty conclusion. For the presence of a suitable identification component in the background presuppositions of a subject's judgment makes for the possibility that in metaphysically and—more importantly—*epistemically* abnormal conditions, the subject's memory judgments *could* be grounded in an identification component. In particular, they might be so grounded if the subject were also informed, *prior* to making his judgment, of the live possibility that he might be storing memory information deriving from someone else's past. Knowledge of the fact that he might be (quasi)-remembering someone else's past, together with grounds for supposing that this live possibility did not in fact obtain, would—in a rational subject—have the effect of *moving* the relevant identities from the role of tacit presuppositions into the role of the subject's own acknowledged rational grounds. Asked to justify his final judgment "I was in that remote part of Scotland five years ago," he could then reasonably offer the following: I seem to remember being in that remote part of Scotland five years ago, and I am, fortunately, the person whose past is responsible for my occurrent seeming memories. In such a scenario, the final judgment would, after all, be liable to error through misidentification relative to the subject's own grounds.

On this proposal Evans would once again be right in holding that the relevant judgments are immune to error through misidentification, because no identity belief features as part of subject's rational grounds for such judgments in normal cases. But Shoemaker would be right at least to the extent that this normal epistemic architecture is a contingency—that once informed that I might be storing memory information deriving from someone else's past, it could no longer rationally be taken for granted that I was not, and my memory judgments would after all rationally require grounding in a fallible identification component.

Note that, even if so, Evans's contention that (apparent) memories have irreducibly first-personal contents would be respected. For, as we saw, the role of the identification component in this case is merely to license the passage from the self-ascription of *seeming to remember* being *F* to the self-ascription that one *was F*. More explicitly, the subject's own grounds for his judgment would be:

(1) I seem to remember being in that remote part of Scotland five years ago.
(2) I am the person whose past I seem to remember now.
(3) I was in that remote part of Scotland five years ago.

Since importing the identification component from the background presuppositions into the grounds does not challenge Evans's point, this second proposed resolution of Shoemaker's and Evans's disagreement respects both their fundamental intuitions.

COROLLARY

A corollary of our findings is that the distinction between *de facto* and *logical* immunity to error through misidentification, originally introduced by Shoemaker,[32] and recently challenged by many theorists,[33] can now be properly characterized and endorsed.

As is familiar, Shoemaker's distinction has been taken to amount to the view that while, for instance, memory-based self-ascriptions are immune to error through misidentification *in this world*, where memory information is stored in the usual way, they would not be so in different metaphysically possible worlds where one is storing information deriving from someone else's past. In contrast, psychological self-ascriptions based on introspection would be immune to error through misidentification in *all possible worlds.* Accordingly, the distinction would be merely *metaphysical*: while certain judgments would be immune to error through misidentification only in this world, others would be so in any metaphysically possible world. But, the question is, how could that be? It is agreed that *in this world* judgments such as those based on occurrent memories are not grounded in any identification component. So how could a mere transfer to a metaphysically possible situation bring about such an error, if that error can arise only when the judgment is grounded in an identification component and there is no such grounding in the relevant kind of case?[34]

---

[32] "Persons and Their Past." It should be mentioned that "Self-reference and Self-awareness," on p. 82, introduced yet another distinction: that between *absolute* and *circumstantial* immunity to error through misidentification: "First-person statements that are immune to error through misidentification [are] those in which 'I' is used 'as subject', [and] could be said to have 'absolute immunity' to error through misidentification. A statement like 'I am facing a table' does not have this sort of immunity, for we can imagine circumstances in which someone might make this statement on the basis of having misidentified someone else (such as, the person he sees in a mirror) as himself. But there will be no possibility of such a misidentification if one makes this statement on the basis of seeing a table in front of one in the ordinary way (without aid of mirrors, etc.); let us say that when made in this way the statement has 'circumstantial immunity' to error through misidentification relative to 'I'." Shoemaker's distinction is drawn on the basis of the subject matter of the judgment, namely, whether it is a psychological or a nonpsychological self-ascription. As such, however, it cannot be sustained. It has in fact become commonplace to hold that, on the one hand, also nonpsychological self-ascriptions can be immune, when made on certain grounds and, on the other, that psychological, inferential self-ascriptions (at least) can be liable to error through misidentification. Thus, it should be replaced by the following: there are *different kinds of grounds* on which *any* kind of self-ascription can be made and, among them, only some would lead to judgments that are immune to error through misidentification. In contrast, *all* kinds of first-personal *judgments*—individuated by their subject matter—can be either liable to, or immune to error through misidentification, depending on the "circumstances," that is, the *grounds*, on which they are made.

[33] I take it that Evans and McDowell do both at least implicitly challenge the *de facto* side when they deny that a mere difference in *metaphysical* conditions would make memory-based self-ascriptions liable to error through misidentification. Campbell,

The previous discussion helps us answer these questions, albeit while giving us reasons to qualify the original characterization of the *de facto*/logical distinction. Judgments, I propose, will be *de facto* immune to error through misidentification if *contingently* true identifications—identifications which may be false in other metaphysically possible worlds—feature as part of their *background presuppositions.* So, while metaphysical conditions will be relevant to the determination of the truth value of such identification components, it is their very presence in the epistemic structure underlying a given judgment which makes for its only *de facto* immunity (either relative to one's grounds or to background presuppositions, depending on one's theoretical preferences).

In contrast, Shoemaker's category of judgments that are (rather unhappily described as) *logically* immune to error through misidentification will be the ones that either have *no* identification component as part of their background presuppositions—such as perceptual

---

"Schizophrenia, the Space of Reasons and Thinking as a Motor Process," and Rovane, *The Bounds of Agency*, pp. 220–32, openly challenge the logical side of it. While I have voiced my perplexities about Campbell's and Rovane's proposals elsewhere (see my "Thought Insertion and Immunity to Error through Misidentification," "What There Really Is to Our Own Notion of the Ownership of a Thought," and "Immunity to Error through Misidentification, the Split between Speaker's and Semantic Reference, and the Real Guarantee"), in the following I will argue that Evans and McDowell can, at most, claim for a revision of that notion in a more epistemologically-oriented direction, whose proper explanation crucially depends on allowing for the existence of identification components as part of a judgment's background presuppositions, contrary to Evans's official position (cf. footnote 11).

[34] This is indeed the kind of complaint Evans and McDowell would raise; but see the previous footnote.

demonstrative judgments like "That thing is approaching very fast"[35]—or else, have an *a priori* true one—such as introspection-based self-ascriptions of occurrent psychological states, which at least some theorists[36] construe as based on identification components like "I = the thinker of this thought." For in the former case there would be no identification component which could give rise to an error of misidentification (either relative to one's grounds or relative to background presuppositions). And, in the latter, no matter what role the identification component played, it could not sensibly be questioned, or called into doubt.[37]

These considerations suffice to show, I think, that Shoemaker's *de facto*/logical distinction, developed along the epistemological lines I have been suggesting, is in good standing. Both its categories can be explained and exemplified. But it remains an open question—and, indeed, one well worth pursuing—whether other routine psychological self-ascriptions, for instance of thoughts or sensations, also have identification components as part of their background presuppositions and whether, in such a case, it is an a priori true one. I personally side with Shoemaker in thinking that they do not, but other theorists, such as Peacocke, have supported this claim and other still, like John Campbell, think that the phenomenon of "thought-insertion" is proof of the fact that such an identification component is not even a priori true. This issue will have to be taken further on another occasion.[38] My central point here has been that merely by "revising and deepening our epistemology"—specifically, by invoking the distinction between grounds and presuppositions—we can more satisfactorily resolve the Shoemaker-Evans dispute, in indeed either of two ways, than any resolution proposed hitherto and—in the course of doing so—shed more general light on immunity to error through misidentification in first-person judgment. And, not surprisingly, none of the considerations we have reviewed requires or even remotely suggests the adoption of a bizarre metaphysics of the self or of a peculiar semantics for "I."

---

[35] This example is in Wright, "Self-knowledge: The Wittgensteinian Legacy," p. 20.

[36] See Peacocke, *Sense and Content*, pp. 109–11; and *Being Known*, p. 275; also Rovane, *The Bounds of Agency*, pp. 220–32. Shoemaker "Self-reference and Self-awareness," p. 89, in contrast, holds that it is a "tautology" that in being aware of a given mental state one is aware that it is instantiated in oneself. Hence, on his view, introspection-based psychological self-ascriptions would be logically immune to error though misidentification because they would not contain any identification component as part of their background presuppositions.

[37] Notice, finally, that if immunity to error through misidentification (either relative to one's grounds or to background presuppositions) could depend on the presence of an a priori true identification component, that would require a parallel revision of the corresponding definitions. Furthermore, it would imply the falsity of Evans's equation between immunity to error through misidentification and identification-freedom.

[38] But see my exchange with Campbell (cf. footnote 8) as well as my "Quelle 'clé pour toutes les mythologies' concernant le soi? Une note sur la question de savoir d'où viennent les illusions de transcendance et comment leur resister," *Philosophie* (2006), forthcoming.

ANNALISA COLIVA

University of Modena

NEW BOOKS: ANTHOLOGIES

ABEL, COREY, and TIMOTHY FULLER, eds.: *The Intellectual Legacy of Michael Oakeshott.* Charlottesville, VA: Imprint Academic, 2005. x, 327 p. Cloth $58.00.

ACHINSTEIN, PETER, ed.: *Scientific Evidence: Philosophical Theories and Applications.* Baltimore: Johns Hopkins, 2005. 286 p. Cloth $49.95.

ADAMSON, PETER, and RICHARD C. TAYLOR, eds.: *The Cambridge Companion to Arabic Philosophy.* New York: Cambridge, 2005. xviii, 448 p. Cloth $70.00, paper $29.99.

AHVEL-RAPPE, SARA, and RACHANA KAMTEKAR, eds.: *The Companion to Socrates.* Malden, MA: Blackwell, 2006. xxiv, 533 p. Cloth $149.95.

ALEXANDER, JEFFREY C., and PHILIP SMITH, eds.: *The Cambridge Companion to Durkheim.* New York: Cambridge, 2005. xvi, 426 p. Cloth $75.00, paper $29.99.

ANSELL PEARSON, KEITH, and DUNCAN LARGE, eds.: *The Nietzsche Reader.* Malden, MA: Blackwell, 2006. xlii, 573 p. Paper $17.95.

ARRIGO, BRUCE A., and CHRISTOPHER R. WILLIAMS, eds.: *Philosophy, Crime, and Criminology.* Urbana: Illinois UP, 2006. 291 p. Cloth $50.00, paper $25.00.

AWODEY, STEVE, and CARSTEN KLEIN, eds.: *Carnap Brought Home: The View from Jena.* Chicago: Open Court, 2004. viii, 387 p. Cloth $49.95, paper $34.95.

AYDEDE, MURAT, ed.: *Pain: New Essays on Its Nature and the Methodology of Its Study.* Cambridge: MIT, 2005. xvii, 423 p. Paper $35.00.

BAGGINI, JULIAN, and JEREMY STANGROOM, eds.: *Great Thinkers A–Z.* New York: Continuum, 2004. vii, 275 p. Paper $14.95.

BAILLEI, HAROLD W., and TIMOTHY K. CASEY, eds.: *Is Human Nature Obsolete?* Cambridge: MIT, 2005. x, 422 p. Paper $27.00.

BALÁZS, ZOLTÁN, and FRANCIS DUNLOP, eds.: *Exploring the World of Human Practice: Readings in and about the Philosophy of Aurel Kolnai.* New York: Central European UP, 2005. x, 343 p. Cloth $47.95.

BALDWIN, T.R., and T.J. SMILEY, eds.: *Studies in the Philosophy of Logic and Knowledge.* New York: Oxford, 2005. 291 p. Cloth $55.00.

BARTSCH, SHADI, and THOMAS BARTSCHERER, eds.: *Erotikon: Essays on Eros, Ancient and Modern.* Chicago: University Press, 2005. 338 p. Cloth $29.00.

BASSHAM, GREGORY, and JERRY L. WALLS, eds.: *The Chronicles of Narnia and Philosophy.* Chicago: Open Court, 2005. xvi, 291 p. Paper $17.95.

BATTALY, HEATHER D., and MICHAEL P. LYNCH, eds.: *Perspectives on the Philosophy of William P. Alston.* Lanham, MD: Rowman and Littlefield, 2005. xiv, 310 p. Cloth $60.00.

BEEBEE, HELEN, and JULIAN DODD, eds.: *Truthmakers: The Contemporary Debate.* New York: Oxford, 2005. 200 p. Cloth $75.00.

BEN-MENAHEM, YEMIMA, ed.: *Hilary Putnam.* New York: Cambridge, 2005. xi, 272 p. Cloth $70.00, paper $25.99.

BENHABIB, SEYLA, and NANCY FRASER, eds.: *Pragmatism, Critique, Judgment: Essays for Richard J. Berstein.* Cambridge: MIT, 2004. xix, 379 p. Paper $32.00.

United States Postal Service

Statement of Ownership, Management, and Circulation

1 Publication Title: Journal of Philosophy

2 Publication Number: 0022-362X

3 Filing Date: 9/20/06

4 Issue Frequency: Monthly

5 Number of Issues Published Annually: 12

6 Annual Subscription Price: Individual: $45.00 Institution: $100.00

7 Complete Mailing Address of Known Office of Publication (Not printer) (Street, city, county, state, and ZIP+4)
The Journal of Philosophy, Columbia University
1150 Amsterdam Avenue, MC4972
New York, NY 10027

Contact Person: Pamela C. Ward

Telephone: 212/666-4419

8 Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
Same

9 Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)

Publisher (Name and complete mailing address)
The Journal of Philosophy, Inc.
1150 Amsterdam Avenue, MC4972
New York, NY 10027

Editor (Name and complete mailing address)
Arthur Danto, President
The Journal of Philosophy, Inc., 1150 Amsterdam Avenue, MC4972
New York, NY 10027

Managing Editor (Name and complete mailing address)
John Smylie, Managing Editor
The Journal of Philosophy, Inc., 1150 Amsterdam Avenue, MC4972, New York, NY 10027

10 Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

Full Name	Complete Mailing Address
The Journal of Philosophy, Inc.	Columbia University
	1150 Amsterdam Avenue
	MC49˜2
	New York, NY 10027

11 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address

12 Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526, October 1999 (See Instructions on Reverse)

13 Publication Title: The Journal of Philosophy

14 Issue Date for Circulation Data Below: September 2006

15 Extent and Nature of Circulation			Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)			3500	3500
b. Paid and/or Requested Circulation	(1)	Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541 (Include advertiser's proof and exchange copies)	700	666
	(2)	Paid In-County Subscriptions Stated on Form 3541 (Include advertiser's proof and exchange copies)	2400	2220
	(3)	Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution		
	(4)	Other Classes Mailed Through the USPS		
c. Total Paid and/or Requested Circulation (Sum of 15b. (1), (2), (3), and (4))		▶	3100	2886
d. Free Distribution by Mail (Samples, complimentary, and other free)	(1)	Outside-County as Stated on Form 3541		
	(2)	In-County as Stated on Form 3541		
	(3)	Other Classes Mailed Through the USPS		
e. Free Distribution Outside the Mail (Carriers or other means)			150	150
f. Total Free Distribution (Sum of 15d. and 15e.)		▶	150	150
g. Total Distribution (Sum of 15c. and 15f)		▶	3250	3036
h. Copies not Distributed			250	464
i. Total (Sum of 15g. and h.)		▶	3500	3500
j. Percent Paid and/or Requested Circulation (15c. divided by 15g. times 100)			95.38	95.06

16 Publication of Statement of Ownership
☒ Publication required. Will be printed in the 9/2006 issue of this publication. ☐ Publication not required.

17 Signature and Title of Editor, Publisher, Business Manager, or Owner
John Smylie, Managing Editor

Date: 9/20/2006

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1 Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.

2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.

3 Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.

4 Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.

5 If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.

6 In item 16, indicate the date of the issue in which this Statement of Ownership will be published.

7 Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form 3526, October 1999 (Reverse)

THE JOURNAL OF PHILOSOPHY

BACK ISSUE SALE

Complete volumes
are now available
at the following discounted prices
in order to reduce our inventory:

Volumes	1 through 98	$10 each
Volumes	99 through 101	$25 each

For more information, contact Pamela Ward,
Business Manager, at pcw1@columbia.edu

Visa, Mastercard, American Express and
PayPal are accepted through our website at
www.journalofphilosophy.org

crítica

REVISTA HISPANOAMERICANA DE FILOSOFÍA

Artículos

ROBERT J. STAINTON
Revenge

Discusiones

STEFANO PREDELLI
The Automatic and the Incomplete.
Remarks on Recanati's Literal Meaning

MANUEL GARCÍA-CARPINTERO
Recanati on the Semantics/Pragmatics Distinction

FRANÇOIS RECANATI
Predelli and Carpintero on Literal Meaning

Notas bibliográficas

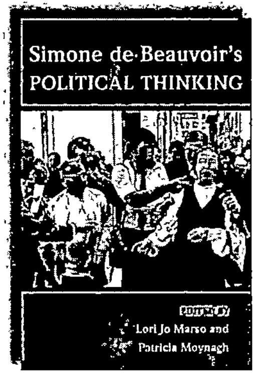
ANITA BURDMAN FEFERMAN Y SALOMÓN FEFERMAN,
Alfred Tarski. Life and Logic [Alejandro Tomasini Bassols]

CHRISTOPHER PEACOCKE, *The Realm of Reason*
[Miguel Ángel Fernández]

Vol. 38 / No. 112 / abril 2006 / ISSN 0011-1503

Crítica, Revista Hispanoamericana de Filosofía aparece tres veces al año en los meses de abril, agosto y diciembre. Es una publicación del Instituto de Investigaciones Filosóficas (UNAM). Todo tipo de correspondencia deberá ser dirigida al Apartado Postal 70-447, Coyoacán, 04510, México, D.F. Correo electrónico: critica@filosoficas.unam.mx.
Versión electrónica: http://critica.filosoficas.unam.mx

UNIVERSITY OF ILLINOIS PRESS



SIMONE de BEAUVOIR'S POLITICAL THINKING
Edited by LORI JO MARSO
and PATRICIA MOYNAGH

Six tightly connected essays home in on the relationship between individual's freedom and one's relationship to community.

Cloth $50.00; Paper $18.00

Recent & Related:

PHILOSOPHICAL WRITINGS
SIMONE de BEAUVOIR
Edited by Margaret A. Simons with Marybeth Timmermann and Mary Beth Mader
Foreword by Sylvie Le Bon de Beauvoir
The Beauvoir Series. Cloth, $40.00

THE COSMOPOLITAN SELF
George Herbert Mead
and Continental Philosophy
MITCHELL ABOULAFIA

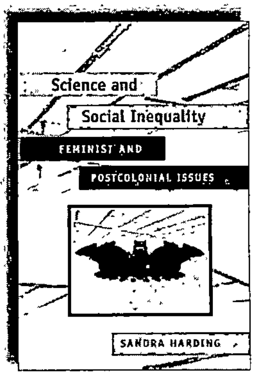
Incisive commentary highlights this survey of the relevance of the American pragmatist philosopher George Herbert Mead to current debates on pluralism, universalism, and the fostering of a democratic temper.

Paper $20.00

KILLING ANIMALS
THE ANIMAL STUDIES GROUP

This collaborative, cross-disciplinary work aims to analyze killing practices while offering a wide academic study of the place of animals in human history.

Illus. Cloth $50.00; Paper $25.00



SCIENCE AND SOCIAL INEQUALITY
Feminist and Postcolonial Issues
SANDRA HARDING

Exposes the ways that hierarchical social formations in modern Western sciences encode antidemocratic principles and practices, particularly in terms of their services to militarism, the impoverishment and alienation of labor, Western expansion, and environmental destruction.

Race and Gender in Science Studies. Cloth $40.00; Paper $20.00

www.press.uillinois.edu

CAMBRIDGE

BEST IN SCHOLARSHIP

Christ and Horrors
The Coherence of Christology
Marilyn McCord Adams
Current Issues in Theology
$85.00: Hb: 0-521-86682-0: 344pp
$29.99: Pb: 0-521-68600-8

Quantified Modal Logic for Philosophers
James W. Garson
$80.00: Hb: 0-521-86367-8: 472pp
$34.99: Pb: 0-521-68229-0

Greek Thought
Christopher Gill
New Surveys in the Classics
$19.99: Pb: 0-199-22074-3: 97pp

Second Edition!
The Emergence of Probability
A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference
Ian Hacking
$70.00: Hb: 0-521-86655-3: 244pp
$24.99: Pb: 0-521-68557-5

Reason's Grief
An Essay on Tragedy and Value
George W. Harris
$75.00: Hb: 0-521-86328-7: 312pp

Doubt and the Demands of Democratic Citizenship
David Hiley
$70.00: Hb: 0-521-86569-7: 200pp
$24.99: Pb: 0-521-68451-X

Aquinas, Aristotle, and the Promise of the Common Good
Mary Keys
$70.00: Hb: 0-521-86473-9: 272pp

Prices subject to change

Arguing about Gods
Graham Oppy
$90.00: Hb: 0-521-86386-4: 472pp

Gandhi's Philosophy and the Quest for Harmony
Anthony J. Parel
$75.00: Hb: 0-521-86715-0: 240pp

The Virtuous Life in Greek Ethics
Edited by Burkhard Reis
$85.00: Hb: 0-521-85937-9: 288pp

Socratic Virtue
Making the Best of the Neither-Good-Nor-Bad
Naomi Reshotko
$85.00: Hb: 0-521-84618-8: 218pp

Darwinism and its Discontents
Michael Ruse
$30.00: Hb: 0-521-82947-X: 328pp

Metaphysics and Method in Plato's *Statesman*
Kenneth Sayre
$75.00: Hb: 0-521-86608-1: 278pp

Human Goodness
Pragmatic Variations on Platonic Themes
Paul Schollmeier
$80.00: Hb: 0-521-86384-8 320pp

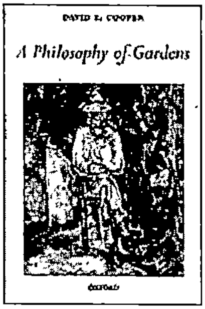
The Heart of Judgment
Practical Wisdom, Neuroscience, and Narrative
Leslie Thiele
$80.00: Hb: 0-521-86444-5: 338pp

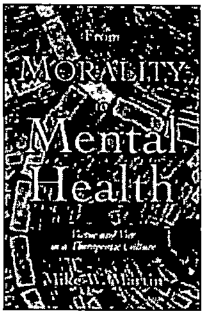
Moral Repair
Reconstructing Moral Relations after Wrongdoing
Margaret Urban Walker
$70.00: Hb: 0-521-81088-4: 264pp
$27.99. Pb: 0-521-00925-1

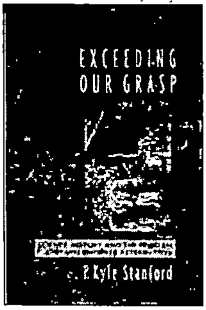
www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Becoming a Subject
Reflections in Philosophy and Psychoanalysis
MARCIA CAVELL

Marcia Cavell draws on philosophy, psychoanalysis, and the sciences of the mind in a fascinating and original investigation of human subjectivity. Cavell brings together for the first time reflections in philosophy, findings in neuroscience, studies in infant development, psychoanalytic theory, and clinical vignettes from her own psychoanalytic practice.

2006 $45 00

A Philosophy of Gardens
DAVID E. COOPER

Given the enthusiasm for gardens in human civilization ancient and modern, Eastern and Western, it is surprising that the subject has been so long neglected by modern philosophy. Now at last there is a philosophy of gardens. David Cooper identifies garden appreciation as a special human phenomenon distinct from both from the appreciation of art and the appreciation of nature.

2006 $29 95

From Morality to Mental Health
Virtue and Vice in a Therapeutic Culture
MIKE W. MARTIN

Morality and mental health are now inseparably linked in our view of character. In this wide-ranging, accessible book, Mike W. Martin asks: are we replacing morality with therapy, in potentially confused and dangerous ways, or are we creatively integrating morality and mental health?

(Practical and Professional Ethics)

2006 $29.95

Exceeding Our Grasp
Science, History, and the Problem of Unconceived Alternatives
P. KYLE STANFORD

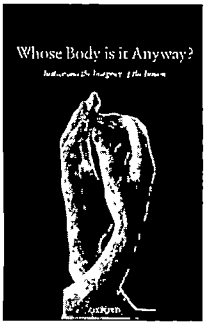
The incredible achievements of modern scientific theories lead most of us to embrace scientific realism: the view that our best theories offer us at least roughly accurate descriptions of otherwise inaccessible parts of the world like genes, atoms, and the big bang. In *Exceeding Our Grasp*, Stanford argues that careful attention to the history of scientific investigation invites a challenge to this view that is not well represented in contemporary debates about the nature of the scientific enterprise.

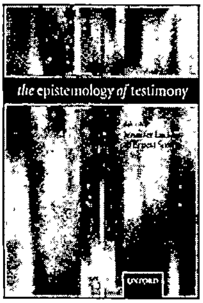
2006 $45.00

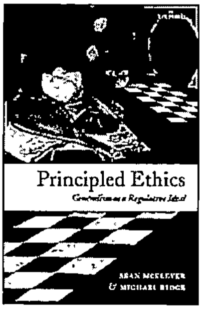
Epistemology Futures
Edited by STEPHEN HETHERINGTON

How might epistemology build upon its past and present, so as to be better in the future? *Epistemology Futures* takes bold steps towards answering that question.

2006 paper $35.00 cloth $99 00







New from
Oxford

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Whose Body Is It Anyway?
Justice and the Integrity of the Person
CÉCILE FABRE

Cécile Fabre's controversial and original book teases out the unexplored implications that arguments for distributive justice have for the rights we have over ourselves, by looking at topical issues such as good Samaritanism, organ confiscation, organ sales, prostitution, and surrogate motherhood.
2006 $55.00

Empiricism and Experience
ANIL GUPTA

"*Empiricism and Experience* develops an original, ambitious, and radical view of how experience, together with a background view, can justify perceptual beliefs."
—Christopher Peacocke, Columbia University
2006 $45.00

The Epistemology of Testimony
Edited by JENNIFER LACKEY and ERNEST SOSA

Testimony is a crucial source of knowledge: we are to a large extent reliant upon what others tell us. It has been the subject of much recent interest in epistemology, and this volume collects twelve original essays on the topic by some of the world's leading philosophers.
2006 paper $34.95 cloth $99 00

Principled Ethics
Generalism As a Regulative Ideal
SEAN MCKEEVER and MICHAEL RIDGE

"Despite the attractions of particularism, most philosophers feel that it misses something. McKeever and Ridge turn these vague worries into razor-sharp objections. Then they develop a moderate alternative view of how principles work in morality. For anyone tempted to become a particularist, this book is a powerful deterrent."—Walter Sinnott-Armstrong, Dartmouth College
2006 paper $35 00 cloth $95.00

Natural Moralities
A Defense of Pluralism
DAVID B. WONG

"This is an exemplary study in the manner of the position it defends: pluralist, empirically grounded, humanistic in the best sense, and widely respectful of difference not just as otherness, but as other ways of being human."
—Barbara Herman, UCLA
2006 $49.95

COMING SOON:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy will be adding
a print and online subscription option for institutions
to begin Fall 2006

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

We expect to launch this new feature with our September 2006 issue.
Please consult our website at www.journalofphilosophy.org
for more details.

University Library
Datta
14.2.07
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CIII, NUMBER 9
SEPTEMBER 2006

page
427 *Conceptual Analysis Naturalized: A Metaphilosophical Case Study* — Christopher Hitchcock
452 *The Demonstrative and Identity Theories of Quotation* — Paul Saka
BOOK REVIEWS
472 Nomy Arpaly: *Unprincipled Virtue: An Inquiry into Moral Agency* — Candace Vogler
478 NEW BOOKS: ANTHOLOGIES

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Libraries and Institutions (print and online)	$180.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.

POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2006 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CIII, NO. 9, SEPTEMBER 2006

CONCEPTUAL ANALYSIS NATURALIZED: A METAPHILOSOPHICAL CASE STUDY*

Once upon a time, possibly for a brief moment in interwar Vienna or postwar Oxford, but more likely in a fairy-tale kingdom, there was a clear understanding of the goal of philosophy. The subject matter of philosophy was a cluster of concepts, typically concepts that are presupposed in science or in other aspects of human endeavor; such concepts include knowledge, causation, freedom, personal identity, goodness, existence, and so on. The goal was to analyze these concepts—to give necessary and sufficient conditions for their application—in terms of concepts that were deemed less problematic (such as empirical regularities). The phrase 'conceptual analysis' is often used to denote this very specific enterprise. But it would not be unreasonable to broaden the definition of 'analysis' to include the kind of careful scrutiny to which philosophers continue to subject concepts like knowledge, freedom, and so on. Conceptual analysis in this broader sense still comprises much of philosophy today. But under this broadened definition of 'analysis', conceptual analysis is no longer the exclusive provenance of philosophy. Empirical psychology attempts to explore the nature of our concepts in a systematic way; indeed, it is hard to imagine how the investigation into the nature of our concepts could fail to have an empirical dimension. What is the relationship between these two very different modes of inquiry? Can the results of empirical psychology inform the philosophical project of conceptual analysis? I attempt to

---

* For comments and suggestions I would like to thank Alex Byrne, David Chalmers, Richard Corry, Clark Glymour, Alison Gopnik, Frank Jackson, Joshua Knobe, Sarah-Jane Leslie, Sarah McGrath, Ruth Millikan, Shaun Nichols, Dan Osherson, Hugh Price, Kim Sterelny, Michael Weisberg, Jim Woodward, and audience members at Australian National University, Florida State University, Princeton University, and the University of Sydney.

0022-362X/06/0309/427–51

explore these issues using my favorite philosophical concept, causation. Of course I will be happy if my explorations make a positive contribution to the understanding of causation, but my main purpose is to treat the concept of causation as a metaphilosophical case study.

In the past few years, there has been a small but growing movement in empirical philosophy. Much philosophical theorizing is built upon our untutored "intuitions" about various hypothetical situations. All too often, these intuitions are simply those of one person, or perhaps a handful of people polled informally. So one natural line of research is to determine whether people do in fact have the sorts of intuitions that are taken to support or condemn various philosophical analyses. Even if we are not willing to accept the intuitions of experimental subjects as probative in philosophical theorizing, psychology might teach us something about the processes whereby intuitions are generated. It would be interesting to know, for example, whether there are strong *framing effects*—whether our intuitions about a case depend upon just how it is presented to us. Or perhaps there are cultural differences in making certain kinds of judgments. For example, E. Machery and collaborators recently found that many of the intuitions that are taken to support direct theories of reference were widely shared among American respondents, but rejected by those from Hong Kong.[1] This gives rise to the possibility that our notion of reference is culturally specific.[2] This use of empirical methods to provide a reality check on the intuitions used to defend philosophical analyses is certainly to be lauded. While I suspect that such an empirical test of the intuitions relied upon in the causation literature might yield some surprising results, this will not be my focus in the present paper. Rather, I am searching for a deeper connection.

There are many different ways in which philosophers try to analyze concepts. As noted above, the canonical approach is to offer necessary and sufficient conditions for the application of a concept in terms of other concepts deemed to be more basic. I will instead focus on a "theoretical role" approach to the analysis of concepts. This approach aims to analyze concepts by articulating their role in facilitating inferences of various kinds. I will argue that the results of empirical psychology can enrich our understanding of the theoretical role of the concept of causation. By the same token, an adequate philosophical account of the theoretical role of the concept of causation can equip us to better interpret a variety of psychological findings.

---

[1] Machery, R. Mallon, S. Nichols, and S. Stich, "Semantics, Cross-cultural Style," *Cognition*, XCII (2004): B1–B12.

[2] M.W. Morris and colleagues suggest that there may be cultural differences in notions of causation, although their discussion does not take place at the level of intuitions about specific cases: Morris, R.E. Nisbett, and K. Peng, "Causal Attribution Across Domains and Cultures," in Dan Sperber, David Premack, and Ann James Premack, eds., *Causal Cognition: A Multi-Disciplinary Debate* (New York: Oxford, 1995), pp. 577–612.

My program is as follows: In section I, I will provide a brief sketch of some of the traditional philosophical approaches to understanding causation. I will not expound upon the theories in any detail, but just rehearse their basic elements. Each account naturally suggests certain views about the epistemology of causation, and about the uses to which causal knowledge is put; I will draw particular attention to these. In section II, I will review the theoretical role approach to analyzing concepts, and offer the beginnings of a theoretical role analysis of causation. In section III, I will outline some recent trends in the empirical study of causal reasoning. In the final sections, I will argue that the theoretical role analysis and the empirical results serve to illuminate one another.

I. THEORIES OF CAUSATION

I.1. Regularity Theories. According to regularity theories of causation, causal relations are underwritten by empirical regularities. In the simplest formulation, associated with David Hume, an event *c* is a cause of a later event *e* just in case events similar to *c* are always followed by events similar to *e*. This simple formulation is universally recognized to be inadequate for a number of reasons. For one thing, there are infinitely many ways in which one event might be similar to, or different from, another. Moreover, many causes fail to be invariably followed by their effects; smoking causes lung cancer, but not all smokers develop lung cancer. These and other difficulties have led regularity theorists to posit more sophisticated patterns of co-occurrence as the foundation of causal relations. According to John Mackie,[3] for example, a property or event-type *C* is said to be an *inus* condition, a kind of minimal cause, for *E* when there is a cluster of properties or types *ABDF*...which sometimes fails to be conjoined with *E*, but is invariably followed by *E* when conjoined with *C*. I also intend for probabilistic theories such as those put forward by Patrick Suppes,[4] Nancy Cartwright,[5] and Ellery Eells[6] to count as regularity theories in a broad sense. These theories do not require that *C* be invariably followed by *E*, even when a complete constellation of additional factors is held fixed: they only require that *E* occur more frequently, or that it be more probable, in the presence of *C*.

---

[3] Mackie, *The Cement of the Universe* (New York: Oxford, 1974).

[4] Suppes, *A Probabilistic Theory of Causality* (Amsterdam: North-Holland, 1970).

[5] Cartwright, "Causal Laws and Effective Strategies," *Noûs*, XIII (1979): 419–37.

[6] Eells, *Probabilistic Causality* (New York: Cambridge, 1991).

Regularity theories suggest that we learn about causal relationships by observing regular patterns of coinstantiation among properties or types of events. This observation may take many forms. It may be completely unconscious, as in the case of classical conditioning in animals. Pavlov's dogs, famously, learned to associate the ringing of a bell with the immanent provision of food. The observation may be conscious and informal, as when we note that a colleague is usually in an irritable mood after meeting with the dean. A number of psychologists have proposed that we learn about causal relationships from discrepancies in the conditional frequencies of observed outcomes.[7] Or we may keep very precise records of our observations and examine the patterns using various statistical methods such as regression and analysis of variance.

Knowledge of regularities is particularly useful in the context of prediction. Knowing that certain antecedents are invariably or frequently attended by particular effects, we can, upon observation of those antecedents, predict the occurrence of those effects.

I.2. Process/Mechanical Theories. Process theories focus on the processes that transmit causal influence across space and time, as well as on the interactions between them. Causal processes, such as rocks, electrons, photons, and sound waves, must be distinguished from mere pseudo-processes such as shadows and spots of light. The most prominent proposal for marking this distinction cites the ability of causal processes to transmit momentum, energy, and other conserved quantities.[8] A pulse of light fired from a laser will carry energy from its source to a point on a wall that is temporarily illuminated. If the source is rotated so that the spot of light moves along the wall, energy will be present at the various points along the wall occupied by the spot. Nonetheless, the spot of light is not *transmitting* the energy, which is instead being freshly supplied by the light source. Process theories are sometimes called *mechanical* theories of causation. They emphasize action by contact and the production of motion, and more generally the importance of the physical mechanism whereby the cause exerts its influence over the effect. I will include in this broad category of causal theory other approaches that are "mechanistic" in the sense that they lay emphasis on the processes and mechanisms that connect cause and effect, rather than on the patterns of correlation or dependence between the cause and effect.[9]

---

[7] See, for example, P. Cheng and L. Novick, "A Probabilistic Contrast Model of Causal Induction," *Journal of Personality and Social Psychology*, LVIII (1990): 545–67.

[8] See especially Wesley Salmon, "Causality without Counterfactuals," *Philosophy of Science*, LXI (1994): 297–312; and Phil Dowe, *Physical Causation* (New York: Cambridge, 2000).

We can acquire knowledge of processes and mechanisms through observation, which can be enhanced by all manner of detection devices (microscopes, Geiger counters, cloud chambers, and so on). Distinguishing causal processes from pseudo-processes requires the application of physics, either naïve or sophisticated. A certain amount of naïve physics may be innate; the famous "launching" experiments of A. Michotte[10] seem to suggest that we directly perceive physically possible collisions to be different from perceptually similar stimuli. Using a dishabituation technique to measure changes in attention, A. Leslie and S. Keeble[11] showed that even infants discriminate between the exchange of motion produced in collisions and other random sequences of moving objects.

Knowledge of causal processes and interactions is particularly useful for understanding the workings of complex systems, such as machines comprised of gears, levers, and pulleys that interact with one another. By way of illustration, here is a common sort of experience that happened to me recently. My toilet would not stop refilling, so I opened up the tank. I had no particular knowledge of the workings of toilets or of plumbing more generally; nonetheless I was able to observe the handle, the lever, the chain, and the ball; and fathom the workings of the device. It turned out that the chain had become tangled, so that the ball could not be lowered far enough to stop completely the hole through which water flows from the tank into the bowl.

I.3. Manipulability Theories. Manipulability theories emphasize that causal relations (in contrast to spurious correlations, for example) can be used by agents to manipulate and control their environment. Causes are "handles" that can be "wiggled" in order to effect changes elsewhere in the world. Critics have charged that manipulation is itself a causal notion, and hence cannot be used to analyze causation. But this does not rule out the possibility of interesting interdefinitions of causation in terms of manipulation, and manipulation in terms of causation, and these interdefinitions may impose a set of highly nontrivial constraints on the possible structures of causal systems.[12]

---

[9] See, for example, the discussions of the notion of mechanism in S. Glennan, "Mechanisms and the Nature of Causation," *Erkenntnis*, XLIV (1996): 49–71, and P. Machamer, L. Darden, and C. Craver, "Thinking about Mechanisms," *Philosophy of Science*, LXVII (2000): 1–25.

[10] Michotte, *The Perception of Causality* (New York: Basic Books, 1963).

[11] Leslie and Keeble, "Do Six-month Old Infants Perceive Causality?" *Cognition*, XXV (1987): 265–88.

[12] The most detailed account is James Woodward's *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford, 2003).

Manipulability theories suggest that we learn about cause and effect through our own interventions on the world. Again, this learning can come at many different levels. At the most basic level is operant conditioning. When an action is regularly followed by a reward or punishment, we adjust our behavior (often unconsciously) so as to perform that action more or less frequently. We also engage in more conscious trial and error learning. At the most sophisticated level, we perform controlled experiments, such as double blind clinical trials. When we wish to test whether a new drug is effective at reducing hypertension, for example, we do not simply allow subjects to choose for themselves whether to take the drug, and then observe the results. Rather, we intervene to determine who takes the drug and who does not.

Manipulability theories focus on our role as agents in the world. Causal knowledge allows us to intervene effectively in the world in order to achieve our ends. This is true at the level of individual decision-making, as well as at the level of national policy.

I.4. Counterfactual Theories. Counterfactual theories attempt to analyze causation in terms of relations of counterfactual dependence. An event *e* counterfactually depends upon event *c* just in case *e* would not have occurred if *c* had not occurred.[13] The simplest counterfactual theory would identify causation with counterfactual dependence. This theory will not work because of problems involving cases of *preemption*, but we need not detain ourselves with the epicycles.[14]

At the most immediate level, we learn about what would have happened under various counterfactual assumptions through a process of imagination or mental simulation similar to that involved in thought experiments. Simulation can also be facilitated by the use of mathematical models or computers. Ultimately, of course, our ability to run such simulations must stem from an understanding of the workings of the world gained from previous experience.[15] Our knowledge about counterfactuals will thus also depend upon the kind of knowledge gained from observation and experimentation described above.

Counterfactual reasoning is particularly apt for retrospective evaluations of responsibility. If one wants to know the extent to which a certain agent's action contributed to an outcome, compare that outcome with what would have happened had the agent behaved differently. This "but for" or "sine qua non" test is often used as a criterion for causal responsibility in the law.

---

[13] I have omitted an explicit statement of the standard qualifiers: *c* and *e* both occur, and they are distinct from one another.

[14] For the classic statement, see David Lewis's "Causation," this JOURNAL, LXX, 17 (October 11, 1973): 556–67.

[15] Ernst Mach famously made this point about thought experiments in *The Science of Mechanics* (London: Open Court, 1883).

II. THEORETICAL ROLE ANALYSES

The sort of approach to the analysis of concepts that I will focus on traces back to the logical empiricists. One classic statement is found in Carl Hempel's paper "The Theoretician's Dilemma."[16] Hempel is concerned with the status of theoretical concepts in scientific theories (although he formulates the problem in terms of theoretical *terms*). He argues that if theoretical concepts can be defined in purely observational terms, then the use of theoretical concepts in science is in principle unnecessary (although theoretical vocabulary may still be useful as a kind of shorthand). However, if theoretical concepts can only be partially interpreted in observational terms (by means of Rudolf Carnap's *reduction sentences*[17]) but this can be done in multiple ways, then theoretical concepts can play a role in what Hempel calls *inductive systematization*. For example (Hempel's),[18] suppose that our theoretical vocabulary contains the term 'white phosphorus', and observational terms 'smells like garlic', 'dissolves in turpentine', and 'ignites at a temperature of 30° C'.[19] 'White phosphorus' is not defined, but is given a partial interpretation:

> White phosphorus smells like garlic.
>
> White phosphorus will dissolve if placed in turpentine.

Our theory will include as a postulate:

> White phosphorus ignites at a temperature of 30° C.

Now if we find that a particular substance smells like garlic, and dissolves in turpentine, then we have inductive evidence that it will ignite at a temperature of 30° C. This inductive relationship is something that we could not capture in a purely observational language. The concept 'white phosphorus' thus plays a certain role in our theorizing, and an understanding of this role provides us with an understanding of the concept. Note that while I talk about the *theoretical* role of a concept, I do not intend to deny that the concept will also be useful in facilitating certain practical judgments. The concept 'white phosphorus' might well play a role in dissuading us from performing a certain procedure requiring temperatures in excess of 30° C.

---

[16] Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 173–226.

[17] See, for example, Carnap, "Testability and Meaning," *Philosophy of Science*, III (1936): 420–68; IV (1937): 1–40.

[18] Hempel, "The Theoretician's Dilemma," in *Aspects of Scientific Explanation*, pp. 214–15.

[19] Unlike Hempel, I here ignore the problem that temperature is also a theoretical concept in need of interpretation.

For another illustration, consider the account of objective chance offered by David Lewis's "Principal Principle."[20] This principle says that:

$$Cr(A|E \,\&\, Ch(A) = x) = x,$$

where Cr is the credence function, or subjective probability, of a rational agent; A is an arbitrary proposition; E is an "admissible" proposition; and Ch is an objective chance function. For simplicity, I have omitted time indices, and I will not pursue the issue of what makes a proposition admissible. In words, the Principal Principle says that a rational agent's degree of belief in A, given that the objective chance of A is x, is equal to x (assuming that the agent is not also conditionalizing upon some inadmissible proposition). This principle does not tell us what kind of thing chance is in the world, but it does tell us a great deal about inferences involving chances. By plugging the principle into a Bayesian framework, it shows us how various pieces of evidence can bear upon hypotheses about the values of chances. Moreover, using some version of expected utility theory, the principle can tell us how information about objective chances should affect our choices. Thus, for example, the Principal Principle can mediate an inference from observed frequencies to the conclusion that one action is to be preferred to another.

My suggestion is that we might try to provide an analogous account of the theoretical role played by the concept of causation, and thereby gain a deeper understanding of that concept. How would such an account go? We may think of the different approaches to causation sketched in section I as identifying different causal *modalities.* (The word 'modalities' here has nothing to do with *alethic* modalities—necessity, possibility, and so on—although these are also obviously connected with causation.) Attempts to analyze causation in terms of these modalities will then be reinterpreted as Carnapian reduction sentences. The role of the concept of causation is to mediate intermodal inferences. To give an oversimplified example in the model of Hempel, suppose that we have the following 2 principles:

Reg: If C causes E in background circumstances B, then in those circumstances E will occur more frequently when C is present than when it is absent.

---

[20] Lewis, "A Subjectivist's Guide to Objective Chance," in Richard Jeffrey, ed., *Studies in Inductive Logic and Probability, Volume II* (Berkeley: California UP, 1980), pp. 263–94; for a similar account, see also Brian Skyrms, *Causal Necessity* (New Haven: Yale, 1980).

> Man: If, in circumstances *B*, intervening to produce *C* is a reliable means for producing *E*, then *C* causes *E* in those circumstances.

Neither is an analysis: Reg states only a necessary condition for *C*'s causing *E*; and Man states only a sufficient condition. We know that the converse of Reg can fail in cases of spurious correlations, and we know that the converse of Man can fail when practical limitations make it impossible for us to intervene so as to produce *C*. Thus there is no logically valid inference from

> In circumstances *B*, *E* occurs more frequently when *C* is present than when it is not

to

> In circumstances *B*, intervening to produce *C* is a reliable means for producing *E*.

Nonetheless the former may still provide inductive evidence for the latter, and the ability to draw such inferences may be extremely useful. Note, moreover, that Reg and Man can serve this purpose even if they are not universally true, that is, even if they do not state genuinely necessary and sufficient conditions, respectively.

For a more sophisticated example, consider the use of Bayes nets to represent causal relationships.[21] I will not attempt a detailed introduction to Bayes nets for the uninitiated, but the basic idea is that one represents a causal system using a graph with arrows connecting variables. The arrows represent relationships of "direct" causation. These relationships are not further defined (except in informal exposition). Rather, there are a set of principles describing the consequences of a causal graph for probabilities, interventions, and counterfactuals. (The connection with causal processes or mechanisms is much less clear.) This permits various kinds of inferences to be drawn: for example, one may be able to infer whether or not a counterfactual is true, or the probability that it is true, on the basis of information about probabilistic correlations and the results of various manipulations. Thus Bayes nets can be used to mediate inferences between these different modalities.

Note that my proposal would still leave a role to play for traditional philosophical analyses. A new and improved regularity theory of causation, for example, would provide a better account of the sorts of inferences that could be made between regularities and causal relationships. But traditional philosophical analysis would no longer occupy a privileged position: developments in statistics and causal modeling could also inform us about inferential relationships between causal relationships and regularities. The principal difference is that philosophical analyses have always aimed at establishing *deductive* inferential connections.

---

[21] See, for example, Judea Pearl, *Causality: Models, Reasoning, and Inference* (New York: Cambridge, 2000); and Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, Prediction, and Search* (Cambridge: MIT, 2000, second edition).

So far, I can only claim to have given a skeleton of an account. My goal in the rest of the paper will not be so much to provide the flesh, as to argue that empirical psychology provides a useful resource for adding flesh to the bones.

III. TWO TRENDS IN PSYCHOLOGY

Like philosophy, psychology has only recently (in the past twenty-five years or so) begun to emerge from a positivistic influence that made causation a taboo subject.[22] Nonetheless, the psychological literature on causal reasoning in humans and other animals is already vast, and I cannot hope to provide a systematic overview of even the little that I do know in a paper such as this. What I will attempt, instead, is to sketch two diverging trends that merit attention. On the one hand, recent developments in developmental psychology indicate that children develop a highly sophisticated capacity for causal reasoning by about the age of four, many pieces of which are already present at a much younger age. By contrast, recent primate studies strongly suggest that even our closest relatives, chimpanzees and other nonhuman primates, lack a wide variety of capacities that appear to require some form of causal understanding. This contrast between young human children and nonhuman primates is particularly interesting since it makes it less likely that the differences in the causal reasoning abilities of humans and other primates is just due to differences in general intelligence. It seems that our capacity for causal reasoning rests neither in our mature scientific understanding, nor in our animal origins. It would take us too far afield to address the issue of whether our capacity for causal reasoning is *innate*, but it does seem to be a basic feature of our uniquely human endowment.

III.1. Causal Reasoning in Primates. Some of the most interesting work to come out of the recent primatology literature has focused on two topics: tool use and social cognition. This is hardly surprising: along with language use, tool use and the complexity of our social arrangements are among the most distinctive features of human culture. Let us begin with tool use. Most of us are familiar with the observations of Jane Goodall[23] and others of chimpanzees stripping the leaves off pliant twigs and inserting them into the channels of termite mounds in order to draw out termites. It is very tempting to attribute to the chimpanzees a great deal of causal understanding on the basis of this behavior. Surely the chimpanzees are selecting pliant twigs so that they can follow the contours of the termite channels as they are inserted; surely the animals are deliberately modifying the shape of the twigs so that they can be inserted easily into the holes. A recent set of very careful experiments conducted by Daniel Povinelli and his collaborators[24] has cast doubt on this interpretation. For example, in one experiment, seven chimpanzees were trained to insert a straight stick through a hole in order to dislodge a piece of fruit that was otherwise inaccessible. The chimpanzees were then presented with a novel tool: a stick that had a cross-piece at one end. The cross-piece was designed to prevent that end of the stick from passing through the hole, and was large enough that its inability to pass through the hole was obvious to humans upon quick inspection. The apes would nonetheless attempt to insert the end of the stick with the cross-piece through the hole, rather than employing the narrow end of the tool.[25] In a related experiment, the chimpanzees were presented a stick containing a removable cross-piece in each end. The animals had previously been given an opportunity to play with these tools, and had learned how to remove and insert the cross-pieces. Instead of simply removing one or both cross-pieces and inserting the modified stick through the hole, the animals would engage in various types of ineffective behavior. In some cases, the animals would remove a cross-piece and then insert it again before trying to dislodge the fruit; in others, they would remove the cross-piece from one end and then attempt to insert the other end through the hole.[26] When the apes did succeed in dislodging the fruit, this success appeared to result from random trial and error rather than deliberate planning.

---

[22] The situation is of course more complicated than this. There has been a strong tradition in psychology, epitomized by the work of Michotte (see note 10), resisting the positivist influence. As noted above, this tradition stresses the apparent observability of causal relations, particularly the causation of motion in collisions between objects. I would argue, however, that this work has nonetheless suffered indirectly from the influence of positivism, particularly from a lack of sympathetic criticism.

---

[23] Goodall, "The Behavior of Free-ranging Chimpanzees in the Gombe Stream Reserve," *Animal Behavior Monographs*, I (1968): 161–311.

[24] Povinelli, with individual chapters co-authored with J.E. Reaux, L.A. Theall, and S. Giambrone, *Folk Physics for Apes: The Chimpanzee's Theory of How the World Works* (New York: Oxford, 2000).

[25] This is, in fact, a rather perfunctory anecdote summarizing the results of three different experiments. See chapter 8 of *Folk Physics for Apes* for details.

[26] See chapter 11 of *Folk Physics for Apes* for a more detailed presentation.

On these and related tasks, a common pattern emerged. When presented with a novel problem, the subjects would gradually learn how to solve it by trial and error, and then internalize a specific procedural rule. When presented with a variant of the original problem, they would initiate a previously learned procedure based upon superficial perceptual characteristics (such as the straightness of the stick) rather than upon the relevant structural features of the problem (the size of the hole in comparison to the end of the stick). This strategy might serve well enough in a natural environment where the superficial perceptual features of a problem are well correlated with the relevant structural features. It is, in fact, quite difficult to distinguish the sort of deliberate behavior that belies an understanding of the causal parameters of a problem from routines learned through trial and error; hence the need for carefully controlled experiments. The limitations of the chimps' strategies were manifest only in a laboratory setting where the perceptual and structural features of a problem could be independently varied.

Another set of results bears upon the much vaunted imitation skills of apes and other primates. Primates do mimic the bodily movements of others, but they seem to lack more sophisticated imitative skills. For example, J. Call and M. Tomasello[27] conducted an experiment in which a stick could be manipulated in a certain way to dislodge a piece of food. The food, and the end of the stick that made contact with it, could not be seen. Orang-utans observed while human demonstrators manipulated the stick to obtain food. When the apes were given their turn, they would manipulate the stick at random, discovering the solution (if at all) by trial and error. The Orang-utans did no better when they observed a conspecific who had mastered the task, rather than a human demonstrator. Whiten and collaborators[28] presented chimpanzees with food in a transparent container that could be opened in one of two ways. One group of chimpanzees was allowed to observe an experimenter open the container in one manner, while a second group observed the container being opened in the other manner. Again, when the chimps had their go, they attacked the box by trial and error. Those in the first group were no more likely to employ the first method of opening the container than those in the second. In each case, the apes seemed to learn an

---

[27] Call and Tomasello, "The Use of Social Information in the Problem-solving of Orang-utans and Human Children," *Journal of Comparative Psychology*, CIX (1995): 308–20.

[28] A. Whiten, D.M. Custance, J.C. Gómez, P. Teixidor, and K.A. Bard, "Imitative Learning of Artificial Fruit Processing in Children and Chimpanzees," *Journal of Comparative Psychology*, CX (1996): 3–14.

affordance of the environment, that food could be obtained in a certain place, and certain objects were brought to their attention, such as the stick in the experiment of Tomasello and Call; but the apes were not able to learn the appropriate means-end relationships by observation. This may be due in part to a deficit in social cognition: the apes could not appreciate that the demonstrators were deliberately employing a known method to obtain food (and similarly in the case where the apes were allowed to observe conspecifics).

We noted above that one context in which we apply causal reasoning is in the assessment of responsibility. It would be interesting to know whether chimpanzees and other primates engage in an analog of responsibility ascribing behavior. I know of no studies addressing this issue in chimpanzees, but there is some evidence that other primates adhere to simple heuristics in situations that would seem to call for responsibility ascription. For example, dominant long-tailed macaques will respond to the cries of a juvenile by indiscriminately attacking subordinates that are closest to the juvenile.[29]

III.2. Causal Reasoning in Human Children. In striking contrast to the results described above, even very young children perform very differently on similar tests. Ann Brown[30] conducted an experiment in which children were given a choice of tools for reaching a toy. The children were allowed to attempt the task as often as necessary to master the skill, receiving help in the form of verbal instructions and pantomime if necessary. The same children were then given the same task with a new set of tools. Even children as young as two years old could select the tool appropriate to the task, in preference to tools superficially resembling (in color, shape, and/or size) the tool originally used to solve the problem. These children, unlike Povinelli's apes, were able to attend to the relevant structural features (such as length and rigidity) of the tools used to solve the original problem, and not merely to perceptual characteristics (such as color and shape). In a similar vein, young children perform very well at imitation tasks. For example, most three and four year olds were able to imitate the actions of a human model in the imitation test of Call and Tomasello described above (*op. cit.*).

Experiments performed by Harris and others[31] show that children are adept at counterfactual reasoning. Children were presented with scenarios acted out with toys. For example, in one scenario, a person walks across a clean surface, getting it dirty. Children are able to reason that the floor would have remained clean if the person's shoes had been removed, and cite the failure to remove the shoes when asked to explain why the floor is dirty.

---

[29] H. Kummer and M. Cords, "Cues of Ownership in Long-Tailed Macaques," *Animal Behavior*, XLII (1991): 357–62.

[30] Brown, "Domain-specific Principles Affect Learning and Transfer in Children," *Cognitive Science*, XIV (1990): 107–33.

[31] P.L. Harris, T. German, and P. Mills, "Children's Use of Counterfactual Thinking in Causal Reasoning," *Cognition*, LXI (1996): 233–59.

Perhaps the most impressive evidence of children's prowess at causal reasoning comes from a series of experiments performed by Alison Gopnik and her collaborators.[32] In one set of experiments, children are exposed to a device called a "blicket detector," and told that blickets "make it go." The children are then shown an object that is identified as a blicket, which is placed on the machine, whereupon the machine plays music and flashes lights. (The machine is in fact operated remotely by the experimenter.) Various objects are put on the machine, either singly or in combination, and children are asked to identify which objects are blickets. Children are able to disregard superficial characteristics of the objects such color and shape; rather, they attend to complex patterns of covariation between the presence of the objects and the operation of the machine in determining which objects are blickets. More specifically, they rely on information about conditional independence relations: if the machine goes off no more often when both objects *A* and *B* are placed on it, than when *A* alone is placed on it, they infer that *B* is not a blicket, even though the presence of *B* is frequently accompanied by the activation of the machine. In one version of the experiment, objects are left on the machine and children are asked to make it stop; children respond by removing the object that had been identified as a blicket.

In a related set of experiments by Laura Schulz, Gopnik, and Clark Glymour,[33] children are shown a machine with a visible crank that turns a gear. A switch can be used to turn the machine on and off. Children are then shown a machine with a switch and two interlocking gears; the mechanism is hidden. When asked to determine which gear(s) is (are) connected to a crank, children will experiment with the switch and remove the gears one at a time to arrive at the correct answer.

---

[32] See, for example, Gopnik and Glymour, "Causal Maps and Bayes Nets: A Cognitive and Computational Account of Theory-formation," in Peter Carruthers, Stephen Stich, and Michael Siegal, eds., *The Cognitive Basis of Science* (New York: Cambridge, 2002), pp. 117–32; Gopnik, Glymour, D.M. Sobel, L.E. Schulz, T. Kushnir, and D. Danks, "A Theory of Causal Learning in Children: Causal Maps and Bayes Nets," *Psychological Review*, CXI (2004): 3–32; Gopnik and Sobel, "Detecting Blickets: How Young Children Use Information about Novel Causal Powers in Categorization and Induction," *Child Development*, LXXI (2000): 1205–22; Gopnik, Sobel, Schulz, and Glymour, "Causal Learning Mechanisms in Very Young Children: Two-, Three-, and Four-Year-Olds Infer Causal Relations from Patterns of Variation and Covariation," *Developmental Psychology*, XXXVII (2001): 620–709.

[33] Schulz, Gopnik, and Glymour, "Preschool Children Learn about Causal Structure from Conditional Interventions," *Developmental Science*, forthcoming.

IV. INTERPRETING THE RESULTS

It appears that there is a whole range of capacities that humans have, even from a very early age, that even our closest relatives do not. Moreover, these capacities are clearly related to causal reasoning. My concern is not so much to *explain* these differential abilities, but to suggest that the theoretical role account of causation sketched in section II above can provide a useful framework for interpreting these results. One important dimension of our capacity for causal reasoning is our ability to draw inferences across modalities such as regularities, mechanisms, and manipulation relations—to use evidence appropriate to one modality in an application appropriate to another. If the account of section II is correct, it is precisely our possession of a full-blown concept of causation that permits us to make these inferences.

The empirical results described in section III.2 establish quite convincingly that human children *are* able to make intermodal inferences. For example, the studies by Gopnik and her collaborators show that children can make use of evidence about regularities to infer the effects of their own actions, and that they can use information about the results of their manipulations to make inferences about unobserved mechanisms. The more problematic claim is that chimpanzees and other nonhuman animals are unable, or substantially less able, to make such intermodal inferences.

First let us note that the (alleged) inability of chimpanzees to make intermodal inferences would be considerably less interesting if they were also unable to make *intra*modal inferences. The astronomical literature on classical and operant conditioning establishes beyond a doubt that mammals are capable of learning about regularities in their environment, even fairly complex ones, and that they are capable of learning about the effects of their actions on the world. To what extent primates are capable of learning about causal processes and interactions is somewhat less clear. Given the sensitivity of their actions to perceptual features of the experimental situation, it is clear that the chimpanzees in Povinelli's experiments could attend to what we might call "geometrical facts," such as the shapes and configurations of objects, whether they are in physical contact with one another, and so on. It is also clear that chimpanzees are able to track moving objects, for example. But it is not clear that this adds up to any kind of understanding of causal processes or mechanisms. What would be desirable, from my point of view, would be something akin to the results of the various launching experiments in human adults and children. These experiments show that humans are able "passively"—without the need for manipulation—to recognize causal interactions between colliding objects. To my knowledge, analogous experiments have not been done with chimpanzees (although Hans Kummer voices the opinion that they would perform similarly to humans[34]). Nonetheless, there have been a number of studies using looking-time paradigms to study primates' mechanical understanding. In one experiment,[35] Laurie Santos and collaborators habituated cotton-top tamarins and rhesus macaques to an L-shaped tool being used to push a grape down a ramp. The primates were then shown two new tools performing the same task—one with the same shape but a different color from the original tool, and a second tool with a shape that made it inappropriate for the task.[36] The primates paid more attention (as evidenced by longer looking time) when the second tool was used, suggesting that this case violated their expectations.[37] Particularly interesting from my perspective are examples of *dissociations* between the performance of primates on looking-time tests and performance based tests. For example, Marc Hauser[38] showed rhesus macaques an arrangement with two boxes, one on top of a solid table, the other beneath it. The arrangement was then hidden behind a screen, and a slice of apple was dropped. The macaques searched for the apple in the box under the table. However, when Santos and Hauser performed a similar experiment,[39] then removed the screen to reveal the location of the apple in the upper or lower box,[40] there was a significant difference in looking times. The macaques looked longer when the apple appeared in the lower box, indicating that this result violated their expectations. This suggests that the macaques understood that the solid table would prevent the apple from passing through, even though they could not make use of this information in guiding their search for the apple.[41]

---

[34] Kummer, "Causal Knowledge in Animals," in Sperber, Premack, and Premack, eds., *Causal Cognition: A Multi-Disciplinary Debate*, pp. 26–36.

[35] Reported in Santos, C. Miller, and M. Hauser, "Representing Tools: How Two Nonhuman Primate Species Distinguish between Functionally Relevant and Irrelevant Features of a Tool," *Animal Cognition*, VI (2003): 269–81.

[36] The grape and the bottom of the tool were occluded by a screen, so the primates saw only the top of the tool moving, and the grape rolling out from behind the screen. The grape was in fact moved by sleight of hand rather than pushed by the tool.

[37] The primates were also presented with a third condition, in which the original tool was used in a different, functionally inappropriate orientation. In this condition, the tamarins exhibited longer looking times, but the macaques did not. This suggests that the macaques may have been attending to the shape of the tool, rather than to its functionally relevant properties.

[38] Hauser, "Searching for Food in the Wild: A Nonhuman Primate's Expectations about Invisible Displacement," *Developmental Science*, IV (2001): 84–93.

[39] Reported in Santos and Hauser, "A Non-human Primate's Understanding of Solidity: Dissociations between Seeing and Acting," *Developmental Science*, V (2002): F1–F7.

[40] The apple slice was in fact placed in the box through a trap door.

What does the empirical evidence say about the abilities of chimpanzees to make intermodal inferences?[42] In the case of the experiments conducted by Povinelli and his collaborators, it strikes me as extremely natural to interpret the results as showing that the apes are unable to make use of the information they have about the shape, configuration, solidity, and so on, of the objects in order to draw conclusions about the effects of their own actions upon them. The dissociation between looking time and searching behavior in rhesus macaques also suggest that these primates are unable to make use of mechanical knowledge to guide their action. According to the interpretive framework that I suggest, this indicates that nonhuman primates lack the concept of causation needed to mediate these inferences.

Povinelli attributes his subjects' poor performance in his tool-using experiments to their positivism. Chimpanzees' attend only to the superficial features of the cases in which their actions succeed and fail: they do not attempt to explain those successes and failures in terms of unobservable causes or concepts such as "force." I find this interpretation somewhat puzzling: *I* would not invoke unobserved causes to explain the inability of the end of the stick with a cross-piece to pass through the hole. The hole and the end of the stick are both observable, and so, presumably, is the "smaller than" relation that holds between them. It does not seem necessary to posit any unobservable entities to explain one's inability to put the wide end of the stick through the hole.[43]

---

[41] Children under two and a half years old exhibit a similar dissociation. Note that I do not attempt to argue that chimpanzees are capable of engaging in the kind of hypothetical reasoning required by counterfactual theories of causation. I do not know how one would test their capacity for such reasoning, although their poor performance on the tool-use tasks set by Povinelli and his collaborators would seem to suggest that the chimps are unable to simulate their own actions in order to anticipate the results. Thus the counterfactual approach to causation is something of an outlier in my picture. One problem here is that counterfactual reasoning is not really intramodal to start with: it must make use of background assumptions acquired in other ways. So the ability to engage in counterfactual reasoning may itself be a manifestation of the capacity to make intermodal inferences.

[42] The following five paragraphs have been strongly influenced by lengthy discussions with Woodward. Woodward makes a number of similar points in section 8 of his "Interventionist Theories of Causation in Psychological Perspective," in Gopnik and Schulz, eds., *Causal Learning: Psychology, Philosophy and Computation* (New York: Oxford, forthcoming).

[43] To be fair, Povinelli seems to have something richer in mind by the notion of "observation" than is standard in the philosophical literature.

Note that the interpretation I favor is slightly different from one that Povinelli considers as an alternative to his own view:

> Perhaps they [chimpanzees] understand specific causal concepts but, for one reason or another, simply can't use them? For instance, perhaps there is only a weak connection between their causal knowledge and their action systems, or they just get stuck in some previous way of responding?[44]

As Povinelli formulates it, this proposal is that the apes have full-blown causal concepts that are not manifest in their manipulations of tools (but which may perhaps be demonstrated through looking-time experiments[45]). My proposal is not that the apes possess full-blown causal information that they are somehow unable to *use*; it is rather that they are unable to translate the broadly mechanical information that they do have into predictions about the effects of their actions; and that this failure is, in part, *constitutive* of their lacking a conception of causation.

What about inferences from information about regularities to the effects of manipulations (or vice versa)? Gopnik, a master of the apt slogan, recently suggested to me that the key question about causal cognition in animals is: "Why don't Pavlov's dogs try to ring the bell?"[46] This captures the spirit of my proposal in a nutshell. Despite the enormous literature on classical conditioning and operant conditioning, there is very little empirical research devoted to intermodal inferences from regularities to manipulations. In one interesting series of experiments,[47] A. Blaisdell and collaborators exposed two groups of rats to two different patterns of association between the sound of a tone and food delivery. One pattern had a temporal order suggesting that the tone was a cause of food delivery, while the other had a temporal order suggesting that the tone and the delivery of food were effects of a common cause (a flash of light). The rats then learned that they could cause the tone by pressing on a lever. The rats that had been exposed to the two different patterns of association differed in the frequency of nose pokes in the food receptacle after pressing the lever, indicating different expectations regarding the appearance of food. Those who had observed the pattern of association suggesting that the tone was a cause of food delivery seemed to exhibit a greater expectancy for food consequent upon pushing the lever. Blaisdell and colleagues interpret this result as showing that the rats are engaging in causal reasoning. Interestingly, however, the two groups of rats did not exhibit any difference in the rate of lever-pressing.[48] Unfortunately, so far as I am aware, no one has attempted parallel experiments on primates. Tomasello and Call express their confidence that chimpanzees would fare no better than rats in the following passage:

---

[44] Povinelli, pp. 322–23.

[45] Povinelli objects that such experiments would not be able to demonstrate the apes' ability to recognize the unobservable phenomena that underlie the causal relations involved in tool use. Again, I find the emphasis on unobservables quite puzzling in this context.

[46] Gopnik, personal communication.

[47] Reported in Blaisdell, K. Sawa, K. Leising, and M. Waldmann, "Causal Reasoning in Rats," *Science*, CCCXI (2006): 1020–22.

> [S]uppose that an individual ape, who has never before observed such an event, for the first time observes the wind blowing a tree such that the fruit falls to the ground. If it understands the causal relations involved, that the movement of the limb is what caused the fruit to fall, it should be able to devise other ways to make the limb move and so make the fruit fall...[W]e believe that most primatologists would be astounded to see the ape, *just on the basis of having observed the wind make fruit fall*, proceed to shake a limb, or pull an attached vine, to create the same movement of the limb...[T]he problem is that the wind is completely independent of the observing individual and so causal analysis would have to proceed without reference to the organism's own behavior and the feedback it might receive from that....[49]

Their hypothetical example involves the observation of a single instance, and not a pattern or regularity, but that seems to be an orthogonal feature; their central claim is that an ape would not be able to infer from what it had passively observed to what it could do to achieve the same end.[50] The results of the imitation tests described above would seem to confirm this diagnosis: they suggest that apes do have difficulties translating the cause and effect sequences they have observed into effective manipulations.

The interpretation that I have been advocating is similar to one offered by James Woodward.[51] Woodward espouses a manipulationist conception of causation, according to which causal relations are those patterns or correlations that persist under (idealized) interventions; such correlations can be exploited for purposes of achieving desirable ends. Spurious correlations, by contrast, are disrupted by interventions, and cannot be exploited for practical ends. Woodward distinguishes three different levels of causal understanding. An agent that exhibits full causal understanding appreciates that the same kind of relationship that she exploits when intervening in the world can exist independently of the actions of any agents. Thus evidence gained by purely passive observation—such as the recognition of correlations or mechanisms—can be used to predict the consequences of her interventions. By contrast, an agent exhibiting purely egocentric understanding recognizes correlations between his actions and various consequences, but does not comprehend that this same relationship among events can exist even when he is not acting. Woodward interprets the failure of apes to solve various causal reasoning tasks as evidence that they have only an egocentric understanding of causal relationships. Woodward suggests that between these two levels of causal understanding is a third, wherein an agent recognizes that the relationship that exists when she intervenes can also be present when other agents interact with the world. This intermediate stage suggests, intriguingly, that our social understanding may play a role in getting us beyond the egocentric stage of causal understanding. Woodward's proposal differs from mine in its emphasis on manipulability relations; on my proposal, full causal understanding requires the ability to make all kinds of intermodal inferences, not merely inferences to and from the effects of one's own actions.

---

[48] Thanks to Gopnik for pointing this out.

[49] Tomasello and Call, *Primate Cognition* (New York: Oxford, 1997), p. 389.

[50] Somewhat surprisingly, Povinelli claims that he and his collaborators would not be surprised to see a chimpanzee behave in the manner described (p. 314).

[51] Woodward, "Interventionist Theories of Causation in Psychological Perspective."

I conclude this section with a cautionary note. Many of the empirical results on which I rely are very recent and/or controversial, and may be overthrown or reinterpreted in light of new evidence. Of course, one cost of adding an empirical dimension to the enterprise of conceptual analysis is that one thereby subjects one's analyses to the risk of empirical refutation. I have presented the research to accentuate the differences between humans and nonhuman primates, but it is not unlikely that the differences may be less dramatic than the experiments just described seem to indicate. There may, for example, be a variety of causal reasoning tasks, some more complex than others, some of which can be solved easily by nonhuman primates, and some which cannot. Hauser and others[52] have found that cotton-top tamarins, which do not employ tools in the wild, could solve a number of causal reasoning problems that were simpler than those facing Povinelli's chimps.[53] For example, when faced with a choice between pulling a blanket that had a food item resting on it, and one with a food item next to it, the tamarins would pull on the former. Or, given a choice between pulling on a blanket that was connected to a piece of blanket that had a food item on it, and pulling on a blanket that was not connected to the one with the food, the tamarins would pull the former. These results suggest that the tamarins have some conception of the principle of action by contact. This is consistent with the results of Povinelli's experiments; for example, the chimps in the experiment requiring the dislodgment of the apple all attempted to use the novel tool to make contact with the piece of fruit on the shelf. Another possibility is that there may be a variety of tasks where apes and other primates are successful at a rate that is significantly greater than chance, but which nonetheless falls short of the sort of reliability with which humans can succeed at the same tasks. It may also be that there are important developmental differences between humans and other primates. Povinelli's experiments were performed on adolescent chimps, while the field studies of T. Matsuzawa[54] suggest that only full adult chimpanzees use tools in the wild. By contrast, human children succeed in causal reasoning tasks involving tool use at a much earlier developmental stage.

---

[52] Hauser, J. Kralik, and C. Botto-Mahan, "Problem Solving and Functional Design Features: Experiments on Cotton-top Tamarins, *Saguinus Oedipus Oedipus*," *Animal Behaviour*, LVII (1999): 565–82.

[53] Thanks to Knobe and Nichols for bringing these studies to my attention.

It is clear that there remains empirical work to be done. A conceptual analysis that motivates interesting avenues of empirical investigation is worthwhile for that reason alone. Even if there is no clean all-or-nothing break between humans and the great apes, the sort of theoretical role account of causation that I have sketched could still provide a useful interpretive framework for understanding the cognitive capacities of different animals. However, since my goal is to argue that empirical results can inform the project of conceptual analysis, I will assume for the sake of argument that further empirical work does indeed support a fairly striking difference in the capacities of humans and other primates to make intermodal inferences, and draw a number of philosophical conclusions.

V. CONCEPTUAL ANALYSIS NATURALIZED

My central thesis is that the sorts of empirical results described in section III can substantially enrich the sort of theoretical role analysis that I have been considering. I do not mean to suggest that one could not formulate such an analysis in the absence of empirical data. Indeed, the discussion of section II suggests that extant accounts of causation can provide ample material for such a theoretical role analysis. My paradigm example of a theoretical role analysis, the Principal Principle, is purely normative in character: it aims to describe how an ideally rational agent would make inferences to and from claims about chances, with no suggestion that actual agents make such inferences. Nor am I suggesting that psychology *supplant* philosophy, that we should let psychology determine what concept of causation we actually deploy and leave it at that. For one thing, we should certainly allow that psychology may reveal ordinary human causal reasoning to be flawed in various respects, but that is only possible if there exist independent means for developing *norms* of causal reasoning. Some psychologists have attempted to develop such norms, but to a philosophical eye they are often very unsophisticated.[55] Moreover, while experimental psychology is well suited to studying the often unconscious causal reasoning used in everyday situations, causal reasoning also occurs in the sciences. In this domain, investigators consciously use specific techniques—experimental methods, statistical analyses, and so on—that will at least sometimes override their common-sense judgments. The examination of these techniques falls within the domain of philosophy of science, not psychology.

---

[54] Matzuwa, "Chimpanzee Intelligence in Nature and in Captivity: Isomorphism of Symbol Use and Tool Use," in W.C. McGrew, L.F. Marchant, and T. Nishida, eds., *Great Ape Societies* (New York: Cambridge, 1996), pp. 196–209.

Nonetheless, there are many important ways in which psychological investigation can contribute to the philosophical project of analyzing a concept such as causation.

V.1. Theoretical Fruitfulness. It is a desideratum of conceptual analyses that they be theoretically fruitful; that they not merely capture the actual or idealized application of the concept, but also illuminate further theoretical endeavors. Thus, it is to be hoped that adequate accounts of freedom and personal identity will shed light on problems in moral philosophy. It is to be hoped that an adequate analysis of the concept of fitness will shed light on issues in evolutionary biology. And so on. I have argued in the previous section that a theoretical role account of causation can shed light on issues in empirical psychology.

V.2. Psychological Realism. Most theories of causation provide criteria (even if not in the form of necessary and sufficient conditions) for judging whether one thing is a cause of another. Psychology can tell us whether the actual procedures we use to judge causation are anything like those proposed by the theory. Such psychological realism is not a *sine qua non* for a successful philosophical analysis; no such analysis purports to be a literal account of how we actually make judgments about causation. Nonetheless, some measure of psychological realism is surely a *desideratum* of analysis—perhaps one that will need to be sacrificed in favor of others, but a desideratum nonetheless. Empirical investigations may be particularly useful in undermining analyses that are bolstered by false but superficially plausible accounts of the psychology of causal reasoning.

---

[55] See, for example, Glymour's critique of A.G. Baker, P. Mercier, F. Vallée-Tourangeau, R. Frank, and M. Pan, "Selective Associations and Causality Judgments," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, XIX (1993): 414–32, in chapters 4 and 5 of *The Mind's Arrows* (Cambridge: MIT, 2001).

The empirical results described in section II.2 suggest human beings, from a very young age, make the sorts of intermodal inferences that would be licensed by the possession of an adequate account of causation. To that extent, the empirical data suggest that the type of theoretical role account of causation sketched in section II above does have a measure of psychological realism.

V.3. Theoretical Niche. As Hempel remarked, if it were possible to reduce theoretical concepts to observational ones, the former would be dispensable to theorizing, at least in principle. It is surely a prerequisite for the success of any theoretical role account that there be a need for something to play the role in question. To use an ecological metaphor, there must be a "theoretical niche" to be filled. The poor performance of apes in making intermodal inferences suggests that these inferences are cognitively nontrivial. This indicates that there is indeed a genuine need for something to play the sort of theoretical role that I have attributed to the concept of causation.

V.4. Expanded Theoretical Role. Although I have focused on the role of the concept of causation in mediating inferences between regularities, mechanisms, and manipulations, the concept of causation will surely play a much larger role within our life projects, and it would seem almost impossible to further articulate this role without leaving the armchair. For example, it is tempting to think that any creature, or at any rate any creature that engages in complex interactions with its social and physical environment, would need a concept of causation just to get by. Thus Michael Waldmann writes: "Our ability to acquire causal knowledge is central for our survival."[56] If recent developments in primatology really do show that even chimpanzees lack anything like our notion of causation, then this must be wrong: it is possible for an organism to fare quite well without engaging in anything like full-blown causal reasoning. The advantages that accrue to those who are able to employ a concept of causation come not at the level of avoiding predation and starvation, but perhaps at the level of tool construction and sophisticated social interaction. Thus the role of specifically causal reasoning in our lives may be more subtle than has previously been appreciated.

---

[56] Waldmann, "Knowledge-based Causal Induction," *The Psychology of Learning and Motivation*, XXXIV (1996): 47–88, on p. 47.

V.5. A Metaphysical Interlude. A natural objection[57] to the line of argument I have been developing is that I am placing too much weight on the term 'concept'. It is handy to have an umbrella term like 'concept' to encompass the various things—causation, freedom, rationality, goodness, knowledge, and so on—that philosophers study. But it is a mistake to think that what philosophers are really interested in is the nature of our concepts. Rather, philosophers are interested in causation, freedom, rationality, goodness, knowledge, and so on.

In reply, I note that it is a common strategy for philosophers to attack their subject matter via our representations of it, be they "Ideas" (Locke and Hume), "Intuitions" (Kant), words (the logical positivists), or what have you. Even if what philosophers are really after is causation itself, it is hard to see how we can identify causation itself without being able to judge its conformity to our concept of causation. Thus conceptual analysis, in the very literal sense of gaining an understanding of the nature of our concepts, is at least a necessary prolegomenon to the analysis of causation, freedom, rationality, and the rest.

There is a fairly natural affinity between the sort of theoretical role analysis of concepts that I have been discussing and the sort of functionalist account that has been dubbed the "Canberra Plan." In the case of causation, such a functionalist account has been offered by Peter Menzies.[58] I have argued that the theoretical role of the concept of causation is to facilitate intermodal inferences; a natural further step would be to claim that causation itself is that feature of the world that renders such inferences legitimate. In other words, causation itself is that feature of the world that is typically present when there are characteristic regularities, which typically makes possible certain interventions, and so on. So far, my presentation has been agnostic on the question of whether there really is something there in the world that corresponds to the concept of causation. As with other theoretical concepts, I think that one could be a realist or an antirealist toward causation, and consistently maintain a theoretical role account of the concept of causation.[59] I think, however, that there is at least some reason to be skeptical of the full-blooded realist position; this will be the subject of the next subsection.

---

[57] Pressed upon me independently by Byrne and McGrath.

[58] Menzies, "Probabilistic Causation and the Pre-emption Problem," *Mind*, cv (1996): 85–117.

[59] I am especially indebted to audience members at the Australian National University—including Byrne, Chalmers, Jackson, Millikan, Sterelny, and Weisberg—for discussion of these issues.

V.6. Humean Modesty. It is worth remembering that Hume himself was as much a psychologist as a philosopher. One way of conceiving of the program of the *Treatise* is as an empirical investigation into the limits of human cognitive capacities. He used the results of this investigation to argue for a kind of modesty about the powers of human reason; we are more like ordinary animals than Aristotle's vaunted rational animals. The reason those tomes of metaphysics and theology contain nothing but sophistry and illusion is that they deal in subjects for which the human mind is just not fit.

Exploring the limits of the human mind is a tricky business: How does one discover limits without being able to see what lies beyond them? Reflecting on the limits of other species can provide an instructive analog. If a chimpanzee were to do philosophy, it seems a safe bet that she would not fare well in attempting to analyze causation. This failure would be due to a fundamental lacuna in her cognitive make-up that she would likely be unable to recognize. We might well be in a similar situation. In evolutionary time, we are mere beginners when it comes to causal reasoning. While nature has given us the capacity to make a heuristically useful family of intermodal inferences, there is no guarantee that we are tracking some specific feature of the world when we do so.

V.7. Self-Aggrandizing Conclusion. Recent developments in primatology suggest that we, meaning members of the species *homo sapiens,* are unique in possessing a full-blown concept of causation. This helps to put the enterprise of analyzing this concept in broader perspective. Philosophy is after all, part of the *human*ities. Nonetheless, there are certainly times, typically when I am constructing ever more convoluted counterexamples involving late preemption by double preventing omission, when I feel as though my work is disconnected from the bigger project of the humanities, to understand who we are, our culture, and our modes of expression. If, however, our concept of causation is part of what distinguishes us as a species, understanding it will indeed help us to better understand just who we are.

CHRISTOPHER HITCHCOCK

California Institute of Technology

THE DEMONSTRATIVE AND IDENTITY THEORIES OF QUOTATION*

Broadly speaking there are two approaches to quotation today, the Demonstrative Theory and the Identity Theory. Both agree that quotation involves picturing, displaying, or ostending—in contrast to the old Name Theory—but they differ on a number of other points. These points are spelled out in this section and they are evaluated one by one in the remainder of the paper. In every case I conclude in favor of the Identity Theory.

The classic version of the Demonstrative Theory, that by Donald Davidson, treats statement (1) as equivalent to (1′).[1]

(1) 'Lobster' is a word.
(1′) The expression type of which *this* is a token is a word. Lobster

Similarly, according to Jonathan Bennett,

> every quotation "*X*" means something like 'The type whose every token resembles, in respects $R_1 \ldots R_n$, *this*: *X*'.[2]

Respects $R_1 \ldots R_n$, are language-relative. In English, the relevant respects for identifying tokens as types involve (for instance) top-bottom orientation ('*u*' versus '*n*'), but not ink-color. Reference to $R_1 \ldots R_n$ provides the Bennett account with descriptive content; the use of "this: *X*" provides it with demonstrative content.

---

* I gratefully thank Susan Haack, Justin Leiber, Doug Patterson, Roger Wertheimer, and an anonymous reviewer for helpful comments.

[1] Davidson, "Quotation," reprinted in *Inquiries into Truth and Interpretation* (New York: Oxford, 1979), pp. 79–92. Other versions can be found in Yitzhak Benbaji, "A Demonstrative Analysis of 'Open Quotation'," *Mind and Language*, XIX (2004): 534–47, "Using Others' Words," *Journal of Philosophical Research*, XXIX (2004): 93–112, and "Who Needs Semantics of Quotation Marks?" *Belgian Journal of Linguistics*, XVII (2005): 27–49; Manuel García-Carpintero, "Ostensive Signs," this JOURNAL, XCI, 5 (May 1994): 253–64, "The *Deferred Ostension* Theory of Quotation," *Noûs*, XXXVIII (2004): 674–92, and "*Double-Duty* Quotation," *Belgian Journal of Linguistics*, XVII (2005): 89–108; Laurence Goldstein, "Quotation of Types and Types of Quotation," *Analysis*, XLIV (1984): 1–6; Barbara Partee, "The Syntax and Semantics of Quotation," in Stephen Anderson and Paul Kiparsky, eds., *Festschrift for Morris Halle* (New York: Holt, Rinehart, and Winston, 1973), pp. 410–18; Arthur Prior, *Objects of Thought* (New York: Oxford, 1971), p. 60; and Ori Simchen, "Quotational Mixing of Use and Mention," *Philosophical Quarterly*, XLIX (1999): 325–36. See also notes 2 and 3.

[2] Bennett, "Quotation," *Noûs*, XXII (1988): 399–418.

0022-362X/06/0309/452–71

A more sophisticated version, due to Herman Cappelen and Ernie Lepore, treats (1) as equivalent to a complicated quantificational structure which they simplify as (1″).[3]

(1″) $\forall x$ (x is a word if x tokens the same as *that*). Lobster

For the sake of brevity, not to mention misgivings about taking (1) as a disguised quantificational structure, I will sometimes represent the Demonstrative analysis as (1*) instead of (1′) or (1″). Nothing hinges on my doing so, however.

(1*) Lobster. This/that is a word.

Several components of the Demonstrative account can be discerned.

(I) The paratactic thesis: quoted matter is logically external to the quoting sentence.
(II) The reference thesis: quotation marks are deictic referring terms, or correspond to quantificational structures that contain deictic referring terms.
(III) The mark thesis: mentioning an expression requires using quotation marks.
(IV) The congruence thesis: quoting is the same as mentioning, and an explication of one is an explication of the other.

These form a tidy package. To begin with, (I) implies (II), which is to say that (I) gives grounds for inferring (II). For if the quoted matter in (1) is removed from the sentence then the remaining portion, (2), will be ungrammatical unless the pair of quotation marks is understood as the grammatical subject, and hence as a referring term or a quantificational structure containing a referring term.

(2) “” is a word.

Conversely, if quotation marks serve as subject then (1) will be ungrammatical unless the quoted matter is regarded as parenthetical or extra-sentential; (II) implies (I). Furthermore, were the quoted matter in (1) actually extra-sentential, and were there no quotation marks to serve as referring terms in the remaining sentence, the remaining sentence would be ungrammatical; (I) implies (III). Finally, given that mentioning invariably aligns with quotation, accounting for one may automatically account for the other; (III) implies (IV).

---

[3] See Cappelen and Lepore, “Varieties of Quotation,” *Mind,* CVI (1997): 29–50; its revised edition, “Semantics of Quotation,” in Kumiko Murasugi and Robert Stainton, eds., *Philosophy and Linguistics* (Boulder CO: Westview, 1999), pp. 209–21; and “Varieties of Quotation Revisited,” *Belgian Journal of Linguistics,* XVII (2005): 51–75.

For some, the Demonstrative package comes apart, though it is not explained how this is possible: Stephen Read and François Recanati hold (I) while rejecting (II), L. Goddard and R. Routley hold (II) without (I), Ben Caplan holds (I) and (II) without (III), and others equivocate.[4] Nonetheless, I take it that (I) and (II) jointly make up the heart of any Demonstrative Theory, that (III) and (IV) are derivative, and that (I–IV) form a prototypical cluster.

In contrast to the Demonstrative Theory, the Identity Theory holds, to coin a cliché, that what you see is what you get: (1) consists of a single sentence in which, more or less, the quoted matter refers to itself.[5] In particular, my own version of the Identity Theory recognizes two intersecting aspects of language: *functional* distinctions, such as that between use and mention, and *formal* distinctions, for example, between plain expressions and those having quotation marks. More precisely, there is one functional distinction between use and nonuse and another between mention and nonmention, and there is a whole variety of formal distinctions drawn by marks of quotation, including pairs of double apostrophes, pairs of single apostrophes, italics, and indented displays (each carrying their own conditions of use). My work grows out of a cognitivist or psychologistic vision of language, as opposed to one that is "realist," referentialist, or platonist. Consequently I avoid appeals to truth conditions, emphasizing instead mental associations, speech act conditions, and pragmatic factors more generally.[6]

---

[4] Read, "Quotation and Reach's Puzzle," *Acta Analytica*, XIX (1997): 9–20; Recanati, "Open Quotation," *Mind*, CX (2001): 637–87; Goddard and Routley, "Use, Mention, and Quotation," *Australasian Journal of Philosophy*, XLIV (1966): 1–49; Caplan, "Quotation and Demonstration," *Philosophical Studies*, CXI (2002): 69–80. For cases of equivocation, see Wittgenstein, *Philosophical Investigations* (New York: Macmillan, 1953), §16; and John Searle, *Intentionality* (New York: Cambridge, 1983), p. 187.

[5] The Identity Theory is propounded by Ken Akiba, "A Unified Theory of Quotation," *Pacific Philosophical Quarterly*, LXXXVI (2005): 161–71; Herbert Clark and Richard Gerrig, "Quotations as Demonstrations," *Language*, LXVI (1990): 764–805; Gottlob Frege, "On Sense and Reference" (1892), in Peter Geach and Max Black, eds., *Translations from the Philosophical Writings of Gottlob Frege* (New York: Blackwell, 1980), pp. 56–78; Marga Reimer, "Quotation Marks: Demonstratives or Demonstration?" *Analysis*, LVI (1996): 131–41; Searle, *Speech Acts* (New York: Cambridge, 1969), §4.1; and Corey Washington, "Quotation," this JOURNAL, LXXXIX, 11 (November 1992): 582–605. See also notes 6 and 7.

[6] For elaboration of this paragraph, see my "Quotational Constructions," *Belgian Journal of Linguistics*, XVII (2005): 187–212, *How to Think about Meaning* (Dordrecht: Springer, forthcoming), and "Quotation and the Use-Mention Distinction," *Mind*, CVII (1998): 113–35.

While my background assumptions lend plausibility to the Identity Theory, they are by no means necessary. Fregeans and others may come to the same conclusions that I do, to wit:

(I′) The integrity thesis: quoted matter is part of the quoting sentence.
(II′) The reference anti-thesis: quotation marks do not refer.
(III′) The plain-mention thesis: it is possible to mention an expression without using quotation marks (although it is not possible to quote without mentioning).
(IV′) The divergence thesis: quoting and mentioning are distinct, and in particular understanding quotation first requires understanding the use/mention distinction.

If quotation marks do not refer, what do they do? According to what I shall call the semantic approach, they express a near-identity function, a function that maps from an expression *x* to the denoting of expression *x*.[7] According to what I shall call the pragmatic approach, quotation marks are comparable to the determinatives of written Egyptian and commas in English; as punctuation marks rather than words, they serve to disambiguate without denoting. One difference between the two approaches is that the pragmatic approach implies (III′), which implies (IV′), whereas the semantic approach implies (III), which implies (IV).

In support of the Identity Theory, especially pragmatic versions thereof, the remainder of the paper argues for (I′) through (IV′) as against (I–IV).

I. QUOTED MATTER BELONGS TO THE SENTENCE

I begin by countering an argument in favor of quotational parataxis and then I give positive arguments against parataxis.

I.1. Rebutting the Paratacticians. What I call the universality argument for extra-sententialism rests on foreign words and animal sounds. We can quote foreign words and animal sounds, in fact any producible sound or image at all, even though they are not part of English; and being no part of English, they are no part of English sentences.[8] Against this view, suppose that spark plugs manufactured in Canada

---

[7] Andrew Botterell and Robert Stainton, "Quotation," *Critica*, xxxvii (2005): 3–33; Mario Gomez-Torrente, "Quotation Revisited," *Philosophical Studies*, cii (2001): 123–53; Susan Haack, "Mentioning Expressions," *Logique et Analyse*, xvii (1974): 277–94; and Paul Pietroski, "Compositional Quotation," in Murasugi and Stainton, eds., *Philosophy and Linguistics*, pp. 245–58.

[8] This argument is inspired by Caplan.

University ALIPORE CAMPUS

are installed in cars that are otherwise completely manufactured in Detroit. We need not fancy an alternate plane of existence, wherein spark plugs really are separate from cars, to judge of a car that it is American-made. Likewise a sentence containing a foreign element, if put together by an anglophone using an anglophone matrix, can be viewed as English without pretending that on some other plane of existence the foreign term stands outside the sentence.

To put the point less allegorically, ordinary beliefs about individuating language are simply mistaken. Few nonlinguists realize that a language is little more than "a dialect with an army"—that there is no linguistically principled distinction between languages and dialects—and few even have the vocabulary for referring to idiolects. I reject the whole question of language as an immaterial object ("Does *x* belong to language *L*'s lexicon?") and would replace it with questions of living language users ("Does *x* belong to speaker *S*'s repertoire?" "For some identifiable population Σ, does *x* belong to the repertoire of most members of Σ?"). Knowing a foreign word, I can say for instance that 'mahatma' means great-souled, or knowing of a word I can say that Gandhi was called 'mahatma', whatever that means. In so doing, I perform a speech act, about the language of others, entirely in my own idiolect. The fact that an anglophone can quote from another language does not make the quoted matter sentence-external.

I turn now to the case against the paratactic thesis.

I.2. An Empty Quotation Argument. The paratactician argues that quoting a word, and saying something about it, is like pointing at an object, and saying something about it: the word quoted is no more a part of the sentence than is the object pointed to. But if this were true then we should be able to remove quoted matter from its context while maintaining syntactic well-formedness. Just as I can say 'that is an apple' in the absence of an apple (I can do so grammatically if not felicitously), so too should (1) be minimally acceptable in the absence of 'Lobster', that is, (3) should be just as good as (4).

(3) "" is a word.
(4) That is a word.

Without context (3), like (4), will fail to express a proposition; but like (4), it should seem like a complete sentence. Yet it does not. Quoted matter is an integral part of any quotational sentence.

My critics have said that (3) is grammatical, but they equivocate. True, (3) is grammatical if it is taken as quoting a very small blank space, and it is also grammatical if it is taken as a case of plain mentioning, that is, as a case where the pair of quotation marks, without itself being quotation-marked, is called a word. However, the grammaticality status of (3) under *those* interpretations is irrelevant. My claim is that (3), not when it is about white space and not when it is about quotation marks but *when it is not about anything*, is ungrammatical.

Others may reply that grammar is deceptive: a sentence can be ill formed on its surface yet well formed at the level of logical form, and in fact (3)'s logical form is admissible because its lonely quotation marks are covert referring terms. But this position—that quotation marks are both overt (visible at surface level in [3]) and covert ([3] lacks a subject on its surface but not in logical form)—strikes me as inconsistent. That aside, there still is a problem. When surface form deviates from logical form, it does so either as a matter of historical contingency or for a general reason. In the former case we should expect other languages to represent quotation very differently from English, namely we should expect quoted matter to be presented as external to the sentences that normally contain them. But I am not aware of any language that does this. On the other hand, if quotational surface form is deceptive for some general reason that applies to all languages, one would hope to have some clue as to what this reason is, or evidence that it at least exists. But again I am not aware of any.

I.3. The Scare-quote Argument. The paratactic thesis is additionally refuted by scare quotes.

(5) Chang is a "bad" boy.
(6) Lee "shamelessly" flirted.
(7) Ishmael floated "above" the other astronaut.

According to the Demonstrative Theory, these are equivalent to:

(5′) Chang is a that boy. bad
(6′) Lee that flirted. shamelessly
(7′) Ishmael floated that the other astronaut. above

To be well formed, quotation marks must be convertible to any syntactic category—they function as adjective in (5), as adverb in (6), as preposition in (7), and so forth. Although English is unusually liberal in allowing words to convert, I do not know of any English lexeme that can so naturally be used as any part of speech at all, and I would be surprised if quotation marks really could. The implausibility of the case is magnified when we consider those languages having quotation marks that generally prohibit conversion. Furthermore, postulating conversion does not work in cases like:

(8) Chang wants to "call upon" Chung.

Because the quoted matter in (8) does not form a syntactic unit, no pro-noun, pro-verb, or pro-form of any kind can take its place. I conclude that quoted matter is an integral part of the sentence that it appears to be in.

The current argument may be said to fail on the grounds that it changes the topic. Our topic, so the charge goes, is metalinguistic citation as illustrated in (1), whereas the counterexamples on hand are cases of scare quotation. However, I do not believe that I am changing the topic. First, marks of metalinguistic citation and marks of scare quotation are typically identical, which means that there is a prima facie reason to assume that a single theory should handle both; in switching from one to the other I am not equivocating. Second, cross-linguistic patterns reinforce the prima facie appearance: in every case I know of, namely in French, German, and Japanese, metalinguistic and scare quotation take identical forms.[9] Third, my earlier work shows that it makes rational sense to see both metalinguistic and scare quote uses of quotation as reflexes of a single underlying linguistic mechanism.[10]

Responding to a related problem involving "mixed" quotation (a mixture of direct and indirect discourse report), García-Carpintero (*op. cit.*, §2) posits an implicit 'as he puts it' right before the quotation in (9).

(9) Quine says that quotation "...has an anomalous feature."

This move, however, yields a logical form that is uninterpretable:

(9′) Quine says that quotation, as he puts it, this. has an anomalous feature.

Although García-Carpintero's exact idea fails, the basic idea demands attention. For one could argue instead that the quotation marks in (9), functioning as a demonstrative, combine with an implicit reference to satisfaction so as to yield (9″).

(9″) Quine says that quotation [satisfies] this [expression]. has an anomalous feature.

By extension, judicious postulations of implicit content could arguably render (5–8) intelligible. For instance, (6) might be construed as:

(6″) Lee flirted [in a manner described by] this [expression]. shamelessly

---

[9] My thanks to Seiichiro Yasuda for verifying the case of Japanese.

[10] See my "Quotational Constructions."

What, however, is the status of implicit content? Is (9″) supposed to give the deep structure of (9)? The truth conditions of (9)? A paraphrase equivalent to (9) for some practical purposes? Since implicit content varies from case to case, ascribing it to the conventional meaning of quotation marks would be ad hoc and implausible. To hold, on the other hand, that implicit content is imaginatively projected based on context fails to explain how (5–9) can be grammatical. As suggested in earlier work, the utterance "I drive a" can combine with pointing at a Studebaker to yield the partially implicit message that I drive a Studebaker; yet this implicit content renders "I drive a" neither grammatical nor—except in charades, interrupted discourse, and false starts—the sort of expression that is ever actually uttered.[11]

II. QUOTATION MARKS DO NOT REFER

Demonstrative Theorists insist that quotation marks are demonstratives or contain demonstratives. It is strange, then, that they neither define 'demonstrative' nor refer the reader to any of the distinct accounts available in the literature.

I take it that demonstratives do more than point. Fingers point but are not demonstratives. A demonstrative is a verbal element with syntactic properties, being neither name nor pure description yet having referential power. It may refer to a singular object ('this, that'), a plurality of objects ('these, those'), a manner ('thus, so'), a property ('such'), or a degree ('yea'); or it may combine with an incomplete phrase, as a determiner does, to yield a referring phrase. In addition, some languages conceivably have demonstrative verbs, prepositions, and conjunctions. But in all cases to be a demonstrative is to function linguistically, and hence to enjoy syntactic powers. Yet this is not how quotation works.

I begin by discrediting a recurrent argument for the reference thesis and then turn to discrediting the reference thesis itself.

II.1. Rebutting an Analogy. Barbara Partee (*op. cit.*, p. 416) reasonably observes that (10a) and (10b) share a common underlying form. After all, they both use demonstratives, they both use mimicry or display, and in both cases the display occurs outside of the sentence.

(10a) Morry stuck out his tongue and went like this: [gesture].
(10b) Morry went like this: [vocal noise].

---

[11] See my "Quotation and the Use-Mention Distinction," p. 119.

But then Partee concludes that quotations such as (10c) share the same form as well.

(10c) Morry went, "[vocal noise]."

The form of her argument—*a* is an instance of *F*, *b* is an instance of *F*, therefore *c* is an instance of *F*—is not deductively valid. Partee might defend the inference by pointing to similarities among *a*, *b*, and *c*, but this move fails. As for surface similarities, (10a) and (10b) *look* like they both have demonstratives and are paratactic, while (10c) does not. As for the purported logical form of (10c), it cannot be presumed to resemble (10a, 10b) without begging the question. The argument is so weak as to merit neglect except that variants of it continue to crop up in the literature.[12]

II.2. Another Empty Quotation Argument. The notion of a referring term is that it can combine with a predicate to form a complete grammatical sentence. But quotation marks do not do this.

(3) "" is a word.

As established in section I.2, *quoted matter is necessary* for making quotation referential. The very same evidence now yields another conclusion, that *quotation marks are not sufficient* for making quotation referential.[13]

II.3. The Redundancy Argument. Recall that, for Demonstrative Theorists, quotation marks are demonstrative referring terms and quoted material is extra-sentential. Therefore, Recanati and Roger Wertheimer observe, Demonstrative Theorists are committed to understanding (11) as (11′):

(11) The word 'emu' is fun.
(11′) The word that is fun. emu

Yet (11) is well formed while its purported equivalent (11′) is not.[14]

One proposed solution treats (11) as equivalent to:

(11″) The word is fun. emu

The idea, evidently, is that quotation marks do not always function as referring terms; sometimes they are just punctuation. However, to add (11″) as one possible reading for (11) does not remove (11′) as one reading admitted by the reference thesis.

---

[12] For example, by García-Carpintero, "Ostensive Signs," p. 260.

[13] This argument is anticipated by Gomez-Torrente, p. 130. I wish to emphasize it, to point out the two distinct conclusions, to consider how paratacticians might respond to at least one of them (section I.2), and to deny Gomez-Torrente's assertion that it extends against the Identity Theory.

[14] Recanati; Wertheimer, "Signs of Speech," Pacific Division APA Meeting (2005).

Bennett (*op. cit.*) acknowledges the equivalence, though he denies that it is a problem: (12a) and (12b) have the same semantic value, he suggests; but that does not preclude them from combining.

(12a) the composer of *Parsifal*
(12b) Richard Wagner
(12c) the composer Richard Wagner

Likewise, by implication, 'the word' actually does combine with 'that' despite the *apparent* ill-formedness of (11′).

This defense is far from convincing, however. In the first place, in drawing a parallel between (11′) and (12c), Bennett ignores their prima facie difference. If they are fundamentally alike, why do they *seem* so different? In the second place, though (12a) and (12b) co-refer, neither is even a partial synonym, analysis, or logical form of the other. Their information values differ, and (12c) is not redundant. Truly redundant appositives would be deviant:

(12d) the composer the composer of *Parsifal*
(12e) the musician the composer of *Parsifal.*

Since (11) is not deviant, quotation marks cannot contain the sense 'the word'. Moreover, it looks like this argument can be extended to other proposed analyses. Given that (13) is acceptable but (13′) is not, quotation marks cannot mean what Davidson says they do.

(13) The expression type of which 'emu' is a token is fun.
(13′) The expression type of which the expression type of which that is a token is fun. emu

Similar examples can be turned against other Demonstrative analyses.

II.4. No Joking Matter. Another problem for the Demonstrative Theory is that in many cases it cannot interpret speech ascriptions.

(14) "You've got to be kidding," I laughed.
(14′) I laughed that. You've got to be kidding

Because 'laugh' is intransitive, (14′) is not completely meaningful. But (14) is completely meaningful. Therefore (14) and (14′) cannot be equivalent.

Demonstratives need not be singular terms. Were they to take the form of deictic adverbs then (14) would mean:

(14″) I laughed thus. You've got to be kidding

The problem here is that 'thus', if it referred, would refer to some manner; yet the displayed item is a sentence, not a manner. Demonstrative Theorists may correctly observe that a given sentence can be mentally associated with a manner (in this case some manner of being uttered), but this move is puzzling. If Demonstrative Theorists are willing to invoke pragmatic associations, why are they not willing to do so earlier in the game, before attributing unobservable logical forms to quotation marks in the first place?

II.5. The Excess-term Argument. Recanati (*op. cit.*, p. 654) observes that quotation marks can be used even in the absence of a matrix sentence, where referring terms should be out of place. Such freestanding quotation is common in novels, for example, Philip Roth's *Portnoy's Complaint*:

(15) "Alex, I want an answer from you..."
"Nuhhh, nuhhh."

According to the Demonstrative Theory, (15) contains dangling referring terms, and is equivalent to:

(15′) This. Alex, I want an answer from you...
This. Nuhhh, nuhhh.

Quotation marks are used in (15) to ascribe words to parties not explicitly identified and to indicate dialogue boundaries. But referring terms do not serve this function, as illustrated by the deviance of (15′). Therefore the items serving this function, the quotation marks, are not referring terms. And if quotation marks are not referring terms in (15), there is no reason to think that they are elsewhere either.

II.6. The Missing-term Argument. Whereas novelists routinely use unascribed dialogue inside quotation marks, which appears to refute the Demonstrative Theory, playwrights routinely use unascribed dialogue without quotation marks, and this too poses a problem for the Demonstrative Theory. Consider, for instance, this passage from Joseph Heller's *We Bombed in New Haven*:

(16) STARKEY Is it [the mission] dangerous?
MAJOR Not for those who survive.

Demonstrative Theorists must make a choice. One option is to recognize normal font as a mark of quotation when set in opposition to capitalized character names. In this case, (16) would fall to the excess-term objection:

(16′) Starkey, this. Is it dangerous?
Major, this. Not for those who survive.

In response to (16′), which if taken literally would be ludicrous, Demonstrative Theorists may posit a pragmatically understood 'says' in each line. But if they are willing to do that, why not posit a pragmatically understood 'says that' in (16)? And if 'that' is but pragmatically understood in (16), font does not express it. Alternatively, Demonstrative Theorists might opt to regard contrastive font as linguistically inessential. In this case, (16) amounts to:

(16″) Starkey. Is it dangerous?
Major. Not for those who survive.

This constitutes plain mentioning—quotation without quotation marks, if you will. Yet Demonstrative Theorists deny the existence of plain mentioning.

III. PLAIN MENTIONING IS LEGITIMATE

The Identity Theory holds that quotation marks disambiguate material that can intrinsically be either used or mentioned. It follows that even unmarked matter can be mentioned, that is, plain mentioning as in (17) is possible.

(17) Cats has four letters.

In saying that (17) is possible, I mean that it is grammatical, though not necessarily as grammatical as (18), or grammatical in the same way.

(18) Cats have four legs.

My claim rather is that the syntactic well-formedness of (17) is comparable to that of (19), though of course they differ in stylistic acceptability.

(19) 'Cats' has four letters.

In my other work I specify how (18) and (19) rely on very different principles of construction. The same difference applies to (18) and (17).

After arguing that plain mentioning is in fact well formed, I defend the thesis from unsound criticisms and then appeal to reflective equilibrium. On this issue depends the plausibility of the Demonstrative Theory.

III.1. The Use Argument. To settle any debate over grammar, it would help to make our ideas of grammar explicit. I propose the following principle: if speakers regularly and unapologetically use construction *C*, then *C* is grammatical. Note first of all that this is intended as a criterion, not a definition or explication; and as criteria generally are, this one is defeasible. (Regularly and unapologetically using construction *C* because one is ordered to at gunpoint, or doing so in fits of religious mania—"speaking in tongues"—does not count as evidence for *C*'s grammaticality.) Note too that the principle does not work the other way. A construction may be grammatical even if no one ever uses it, for it may be obviously false, too long for human breath, rude or taboo, and so on.

Now the evidence strikes me as perfectly straightforward. First, plain mentioning is widespread. Proof of this, in the form of pages and pages of empirical attestations, would take too much space to enumerate. Second, plain mentioning is resistant to change. After telling my students that they absolutely must use quotation marks when citing words, I invariably get some who fail to do so. (When questioned, they will sometimes say that quotation marks were inappropriate because what was meant was obvious.) Third, plain mentioning can be found in expertly written, professionally copy-edited texts. Lynn Truss, an authority on punctuation, writes of Woody Allen's character:

(20) He just wanted to know how to spell Connecticut.[15]

And Paul Dickson conforms to the practice found in books put out by many fine publishers when he writes:

(21) Chang is the most common name in the world.[16]

Indeed, when context enables the listener to expect a name to be mentioned rather than used, quotation marks are not merely omissible; they are positively deviant.

(22) My name is 'Paul'.

All of this is evidence that plain mentioning is no performance error. The product of competence, it is grammatical.

In addition, it is widely agreed that in speech one can mention a word without vocally marking it in any special way, that is, one can plain mention it. It follows, to the extent that written language is a record of spoken language, that written language may also be plain mentioned yet grammatical.

It has been suggested to me that, though grammatical, (20) is not really a case of plain mentioning because (20) contains a silent demonstrative. I cannot refute this idea because it is unfalsifiable. Suffice it to say that there is no independent phonological, syntactic, diachronic, psycholinguistic, or other reason to suspect a null demonstrative here, and that it is not invoked by any published defense of the Demonstrative Theory.

---

[15] Truss, *Eats, Shoots and Leaves* (New York: Gotham Press, 2004).

[16] Dickson, *What's in a Name?* (Springfield, MA: Merriam-Webster, 1996).

III.2. A Question-begging Criticism. Cappelen and Lepore deny that plain mentioning is possible, giving an argument seconded by Stefano Predelli. Recall that (17) is 'Cats has four letters'.

> Consider a language *E**... *E** has no mention-quotation distinction, and so [17] is ungrammatical in *E**. Ask yourself: could a speaker assertively utter a token of [17], and her audience still understand what she means by her token? Should we expect that speakers of *E** might occasionally omit...quote marks? The answer is trivially "yes." *E** speakers could omit quote marks whenever what's meant is obvious.... If *E** speakers can omit quote marks and still get across what they mean, then the fact that we omit quote marks is no evidence that we are not ourselves *E** speakers....[17]

On the face of it, Cappelen and Lepore contradict themselves. They posit (a) that language *E** has no mention-quotation distinction, that is, its speakers invariably quotation-mark their mentionings; and at the same time (b) that *E**'s speakers occasionally omit quotation marks when mentioning.

Perhaps Cappelen and Lepore mean (a′) that *E** has no *grammatical* mention-quotation distinction, that is, its speakers invariably quotation-mark their mentionings when speaking grammatically; and (b′) *E**'s speakers may omit quotation marks when mentioning as long as "what's meant is obvious," although doing so is ungrammatical. But they err in thinking that obviousness licenses ungrammatical utterances. Whenever a grammatical subject in English denotes the established topic of a conversation, what is meant, however obvious, is not systematically omitted outside of marked genres such as notes to oneself. Also, telegraphic language (not to be confused with elliptical language, which is perfectly grammatical) is usually clear in meaning, and certainly saves on breath, and yet telegraphic speech is generally restricted to telegrams, headlines, aphasic speech, and the like. In short, principle (b′) does not describe English; if plain mentioning is ungrammatical in *E** then *E** is clearly distinguishable from English, where plain mentioning is widely, regularly, and unapologetically used.

More fundamentally, in stipulating a language that has no mention-quotation distinction, Cappelen and Lepore invoke a premise that I am not willing to grant. Quotation marks are forms that a language may or may not possess. Mentioning is a function; it is what people do with language, not a feature of the language itself. In speaking or writing *any* language whatsoever, therefore, I can choose to mention one of its expressions regardless of whether I use a quotational form. There is no such thing as a possible language which lacks the mention-quotation distinction, and in appealing to it Cappelen and Lepore beg the question.

---

[17] Cappelen and Lepore, "Using, Mentioning, and Quoting: A Reply to Saka," *Mind*, CVIII (1999): 741–50; Predelli, "Scare Quotes and Their Relation to Other Semantic Issues," *Linguistics and Philosophy*, XXVI (2003): 1–28.

III.3. The Ambiguity Criticism. If plain mentioning is possible then (23) is ambiguous, allowing for an objectual reading (the font known as Arial is elegant) and a metalinguistic reading (the name of the font is elegant).

(23) Arial is elegant.

Demonstrative Theorists criticize this ambiguity view, treating it as absurd.[18] I do not see why they feel this way unless they think all ambiguity is lexico-syntactic. But such a restricted view of ambiguity matches neither that expressed by the Identity Theory nor that of ordinary language. (Mona Lisa's smile is ambiguous without being lexically or syntactically so.)

Not incidentally, Demonstrative Theorists must themselves posit systematic ambiguity, given that they treat (24) and (24′) as equivalent.

(24) "You are either with us or against us" is an Orwellian lie.
(24′) (a) You are either with us or against us. (b) That is an Orwellian lie.

Apparently (24) is true while (24′) is self-contradictory—(a) contradicts (b). If (24′) is not to be self-contradictory, (a) must lack assertoric force. Yet in ordinary contexts (a) possesses assertoric force. Since some tokens of (a) are assertoric and some are not, (a) is ambiguous.

III.4. Reflective Equilibrium. I judge plain mentioning to be grammatically well formed while Demonstrative Theorists do not. To decide between us, you should look at the theoretical considerations for supporting one judgment over the other. But you should look further, too. You should look for the stories each side can give to explain the intuitions of the other.

Here is mine. As a general rule, if you know an effective means of doing something that you value, you will frown upon those who willfully refuse to adopt it, you will regard their choices as erroneous. Using quotation marks is an effective means of being explicitly precise, which is of value to scholars; hence it is easy for scholars to condemn language that does not use it. The fact that plain mentioning might be eschewed by certain speakers is no evidence of its ungrammaticality.

---

[18] Cappelen and Lepore's reply to Saka, p. 744; Caplan.

Is it credible that smart people can be confused by the distinction between ungrammaticality on the one hand and stylistic infelicity, flouting of scholarly practice, and other sorts of ill-advisedness on the other? Our long sorry history proves that it is so. Grammarians rail against 'ain't', denying that it is a word, when in fact its real problem is prudential. Speakers who use the word acquire lower social status; they do not violate grammar.

To summarize, some regard (17, 20, 21) as well formed and some do not. The best theory not only justifies one side but also explains the existence of divergent intuitions. On my account, those who accept plain mentioning attend to grammar while those who reject them are distracted by other felicity conditions. So far as I know, Demonstrative Theorists cannot account for divergent intuitions. They cannot say, for instance, that those who reject (17, 20, 21) attend to grammar while those who accept them do so because some ungrammatical constructions are assertable. For as a general rule, ungrammatical constructions are *not* regularly and unapologetically assertable.

III.5. The Significance of Plain Mentioning. If thesis (III′) is right then it serves as additional argument against the paratactic and reference theses (I, II). For given that plain mentioning is possible, (19) would keep its meaning even if stripped of its quotation marks, setting aside differences in ambiguity.

(19) 'Cats' has four letters.

But in that case, for purposes of grammaticality and interpretability, quotation marks need not add content to the sentences that they are in (modulo ambiguity). Hence so-called semantic accounts of quotation, including the Demonstrative Theory, are at best superfluous. I conclude that Demonstrative Theorists must do more than show that their account is possible. They must show that simpler accounts do not work. They must show both that it is possible to construe quotation marks as demonstratives *and* that it is not possible to construe them as simpler nonreferring elements.

IV. THE USE-MENTION DISTINCTION IS RELEVANT

Understanding quotation calls for understanding the use-mention distinction. Neglecting the distinction keeps us in the dark, I shall argue; and, furthermore, respecting the distinction is illuminating.

IV.1. Against Neglecting the Distinction. Recall that the Demonstrative Theory treats (24) and (24′) as equivalent, yet (24′) appears to be self-contradictory.

(24) "You are either with us or against us" is an Orwellian lie.
(24′) (a) You are either with us or against us. (b) That is an Orwellian lie.

Davidson responds that (a) is not asserted, it is merely "said." What does it mean to say something? Davidson seems to suggest that saying is nonintentional, that it amounts to uttering, when he defines quotation as a reference to an expression *shape*. The problem is that the objects of utterance, mere syllables, are neither true nor false. If quotation is defined by reference to uttering then (24) claims that sounds, as such, can be false, when in fact it seems that only interpretations of sounds, or sounds relativized to interpretations, have truth value. Davidson himself appears to acknowledge this elsewhere when he says that truth values are borne by structural descriptions. On the other hand, if saying is more than flapping one's gums, to say is to make an assertion, pose a question, issue a directive, exclaim, attribute words to others, cite, or do something meaningful. But if one explains quotation in terms of saying and saying in terms of attributing or citing, one does not get very far. The resulting account of quotation is circular or at best incomplete.

Cappelen and Lepore do spend one sentence on mentioning: "expression *e* is mentioned in an utterance *u* just in case the token of *e* occurring in *u* is produced in order to be demonstrated so as to talk about tokens that same-token it."[19] Not only does this formulation contain minor flaws—for example, speakers talk, expressions do not, and even then speakers do not always talk about the expressions that they mention—more to the point, in consigning their discussion to a footnote Cappelen and Lepore marginalize and obscure the essential relation between quoting and mentioning.

IV.2. For Respecting the Distinction. I believe that by examining the use-mention distinction, specifically how it develops both historically and in the individual, we find evidence for the correct theory of quotation.

I treat the use-mention distinction as a matter of how language is used. This distinction, being independent of the lexico-grammatical peculiarities of specific languages, presumably applies to all languages; a speaker of any language can intend to draw attention to an expression. This is true even for a language without quotation marks.

---

[19] Cappelen and Lepore, "Varieties of Quotation," note 8.

Such is easy to imagine, for ancient languages lack quotation marks, as do juvenile versions of modern languages.[20]

Now imagine that you, a child, wish to write that you have learned a new word one day, 'uranium'. You start by writing, "I learned a new word today." How do you finish?

(25a) It is this. uranium
(25b) It is this: uranium.
(25c) It is this, uranium.
(25d) It is URANIUM.
(25e) It is uranium.

I grant that (25a) and (25b) are real possibilities. This shows that it is possible to mention by means of paratactic demonstrating, though it does not show that such is either necessary or typical, nor that this is how quotation works. A more common and natural way for children to write is probably (25c), which, though demonstrative, appears not to be extra-sententially paratactic. Also to be expected would be (25d). Those children who underline, box, capitalize, or otherwise fancify the quoted matter, as in (25d), are inventing their own marks of quotation, perhaps though not necessarily as an effort to represent in writing some feature of speech. It is the naturalness of the likes of (25e) that establishes that proto-language, be it juvenile English or Old English, does not grammatically require either demonstrative reference to that which is mentioned nor paratactic separation of that which mentions from that which is mentioned.

The question now is this: In the development from ancient to modern language, or from juvenile language to mature, which hypothesis is most plausible for explaining current conventions of quotation?

(α) Elite speakers (editors and whatnot), realizing the ambiguity of (25e), first took (25e) and secondly supplemented it with innovated disambiguating punctuation.
(β) Speakers first took (25a), secondly replaced the ordinary demonstrative with quotation marks, thirdly moved the quoted matter into the sentence at surface form but not logical form, and fourthly banished (25e) as a way of mentioning—and all in perfect synchrony.

The answer, of course, is that simplicity overwhelmingly favors (α).

IV.3. Summary. The Identity Theory, whereby quotation marks signal disambiguating intent, makes several predictions. First, slapdash writers oblivious to possible ambiguity will frequently omit quotation marks. Second, quotation marks will be omitted even by careful editors in cases, such as written plays, where ambiguity is systematically not a problem. Third, if quotation marks were added by some amateur playwright, as in (16*), they would be regarded as inelegant by those familiar with the norms of playwrighting, but not as uninterpretable.

---

[20] Thus, Chrysippus in Brad Inwood and L.P. Gerson, eds., *Hellenistic Philosophy* (Indianapolis: Hackett, 1988), p. 78: "If you say something, this comes out of your mouth; but you say wagon; therefore, a wagon comes out of your mouth."

(16*) STARKEY: "Is it dangerous?"
MAJOR: "Not for those who survive."

Fourth, as first noticed by Wertheimer, it will be unusual to find quotation marks around nonlinguistic material, for lack of ambiguity.[21]

(26) The color within the square, □, is white.

These facts and others favor the Identity Theory over the Demonstrative Theory.

Most crucially, the ill-formedness of example (3) establishes the integrity thesis, that quotations include their quoted matter, and it also establishes the reference anti-thesis, that quotation marks are not referring terms.

(3) "" is a word.

Sections I and II mount additional evidence for the same conclusions, and they furthermore support each other: if either section is correct, that would count as support for the other.

Section III observes that (21) is not the result of performance error, it is grammatical.

(21) Chang is the most common name in the world.

Furthermore, if (21) is grammatical then we seem to have a word being used in reference to itself, in accordance with the pragmatic version of the Identity Theory. We certainly do not have demonstratives doing the referring. All of this suggests that the simplest account of (27) is that it too has no demonstrative referring terms.

(27) 'Chang' is the most common name in the world.

This conclusion is reinforced in section IV, which speculates on the transition from a primitive language to one having quotation.

Cappelen and Lepore write that because their theory is semantic and mine is pragmatic, the two cannot possibly conflict. Were this true then their arguments in favor of one theory over the other would be off the mark. The fact is, pragmatic and semantic explanations can rival each other much as psychiatric and demonological explanations of eccentric behavior do. Indeed, my whole program is to diminish the role of semantics, as understood by platonists and other realists, in all linguistic explanation. Be that as it may, I conclude that the Identity Theory is a contender that deserves consideration, and I demur when Cappelen and Lepore characterize their own account as "enormously attractive," "almost irresistible," and "ingenious."[22]

---

[21] Wertheimer, "Quotation Apposition," *Philosophical Quarterly*, XLIX (1999): 514–19.

PAUL SAKA

University of Houston

---

[22] Cappelen and Lepore, "Semantics of Quotation," p. 217f.

BOOK REVIEWS

Unprincipled Virtue: An Inquiry into Moral Agency. NOMY ARPALY. New York: Oxford University Press, 2003. viii + 203 p. Cloth $55.00, paper $17.95.

Nomy Arpaly's principal thesis in *Unprincipled Virtue* is deceptively straightforward: an action is morally praiseworthy just in case the agent did the right thing for the very reasons that made the action right, and, *ceteris paribus,* the deeper the moral concern informing an action the more praiseworthy the act. By a similar token, moral blameworthiness attaches to actions done for "sinister" (deeply immoral) reasons, and acts done from indifference to, or lack of concern for, moral considerations. On this view agents can perform right actions for wrong reasons or wrong actions by accident (that is, in spite of deep, appropriately targeted concern for the morally salient aspects of their circumstances), and not all agents who succeed in performing right actions for right reasons are equally praiseworthy. While the slender wedge thereby driven between the rightness of an action and its moral worth may run contrary to some work in contemporary ethics, most of us will accept some distinction between an action done in accordance with morality, duty, or virtue (a right action done in a way, at a time, or from a consideration that is not quite right) and an action done *from* morality, duty, or virtue, so this aspect of Arpaly's project ought not to cause her fellow ethicists alarm. Her deep challenges to other work in the field come in her detailed accounts of moral responsiveness and depth of moral concern.

Both moral concern and responsiveness to moral considerations are, on Arpaly's account, *de re* rather than *de dicto.* That is, an agent need not be able to give an accurate account of her reasons, of the morally significant aspects of her circumstances, or of the moral desirability of her action in order to act for the right reasons from deep moral concern. Arpaly argues that an agent need not even take conscious—reflectively endorsed, or merely avowed—interest in doing the right thing in order to be appropriately responsive to moral considerations. By the same token, an agent's avid, nuanced, delicately articulated, reflectively endorsed concern for morality (developed with all due sobriety under conditions conducive to cool, clear-headed rational deliberation) need not issue in morally praiseworthy action. It need not even issue in morally *desirable* (right)

0022-362X/06/0309/472–77

action. Quite the contrary, some agents will be morally better off if occasionally they are driven to act against their own best judgment for reasons that they take to be selfish, freakish, petty, morally insignificant, or outright wicked and shameful. As Arpaly puts it, distinguishing her view from standard, mainstream interpretations of Kant's insistence that only actions done *from* duty have moral worth, or of Aristotle's insistence that the virtuous agent performs fine actions *for the sake of the fine*, "moral worth is fundamentally about acting for moral reasons, not about acting for reasons believed or known to be such, and distinguishing the two is important in evaluating moral agents" (73).

The "virtue" at issue in Arpaly's title, then, is the deep concern and directed, effective responsiveness to moral considerations that make for a good will and that inform morally praiseworthy actions. That virtue on this account is "unprincipled" does not mean it is not a proper *source* of the actions and psychological states at issue in virtuous conduct. In the most traditional sense of "principle"—where the principle of something actual is its ratio and source—Arpaly's virtuous agent is thoroughly principled. Neither does the fact that this virtue is "unprincipled" mean that it is *unsystematic* (to shift to a newer sense of "unprincipled"). For, although virtue need not produce a reliable stream of morally praiseworthy actions throughout the course of an agent's maturity, Arpaly leaves open the possibility that, in tracking the morally salient aspects of situations and circumstances, moral responsiveness and moral concern will take their direction and force from an orderly field of practical considerations (however hard it may be in practice to determine which of those considerations takes precedence when two or more conflict). The virtuous agent's sense of what she is doing and why may be inarticulate and slightly chaotic. What she is up to, and why she merits our praise, need not be. Rather, virtue is "unprincipled" because it need not involve the agent's articulate grasp of the considerations that inform her action and provide foci for her concern when she acts well. The book is devoted to developing and motivating this thought, producing accounts of moral worth and moral responsibility that answer to it, and, finally, defending these against potential objections.

Arpaly insists that her project is primarily positive rather than critical. But her account of moral worth *de re* is at odds with much of the work in several fields, and readers who have greeted the book as a treasure trove of counterexamples to various views about moral motivation, practical reason, virtue, action theory, agency, rationality, autonomy, and the like are not *wrong* to sense formidable challenges to current orthodoxies on the move in this work. Moreover, the acuity of Arpaly's critical edge is honed by her quiet, steady, and fearless insistence that philosophical moral psychologists pay attention to first-order moral psychology in all of its messy detail. To this end, she works by assembling a stunning array of examples—some taken from works of fiction or film, some drawn from life, and some conjured from her fertile moral imagination. A few examples are given extensive treatment. Many arise in passing. All in all, the examples are so good that one might miss the care with which Arpaly positions her own work in the overlapping philosophical regions that she means to engage.

She begins by laying out a series of "types of individuals whose moral complexities have not be the subject of significant discussion" in the relevant philosophical literatures. Her project takes its force from the need to account for these kinds of cases. The way to do so, she argues, is by noticing that, while philosophers interested in rationality, morality, or virtue tend to emphasize agents' self-understanding, moral common sense is much less worried about the first-person perspective. Arpaly takes the prevailing philosophical orthodoxies to be artifacts of reliance upon artificially simplified examples. What we need, she claims, is to attend to complex cases in which the complexity is *itself* moral psychological rather than merely situational.

She considers her cases from the perspective of standard philosophical accounts of rationality in her second chapter, and moves to the heart of her project—her account of moral worth and of moral responsibility—in the third. The clearest philosophical casualty in her account of these matters is standard views about the nature and importance of autonomy in discussions of moral worth and moral responsibility. By the lights of many views, she seems to insist that morally praiseworthy action need not be, in any interesting sense, autonomous action. She begins to address concerns about autonomy in chapter 4. There, she discusses eight distinct senses of "autonomy" at work in contemporary philosophy. Throughout, she deploys an additional battery of examples in the service of arguing that autonomy (in *any* of the featured eight senses of the term) is not a constitutive feature of moral responsibility (although having those higher faculties that are a *precondition* for the development of agent-autonomy—self-governance or self-control—is likely *also* a prerequisite to moral responsibility). The upshot of this chapter is that agent-autonomy is not a necessary condition for moral responsibility. Arpaly devotes her final chapter to addressing an illuminating series of objections that might be raised against just this point. The work of the fourth and final chapters is twofold. On the one hand, Arpaly makes a series of argued interventions into contemporary work on autonomy, morality, and practical rationality. On the other, she continues to display how philosophical work might begin to change if we took seriously moral complexity. Her specific objections to other philosophers' accounts of autonomy are often deep and interesting. But it is her focus on the complexity of moral life that will likely have the biggest impact.

Common-sense morality can countenance points that find no place in contemporary ethics and action theory, Arpaly argues. These points include the thought that people sometimes act well from unconscious motives (and, worse, that people sometimes need to *decrease* their normal powers of self-control if they are to get themselves into a position to act well), that being a good person is not the same as having a knack for abstract reasoning, that cool-headed deliberation is neither a necessary nor a sufficient precursor to right action, that the fact that some desire accords well with one's sense of oneself or of what matters in life does not mean that action in the service of this desire will be in *any* sense praiseworthy, that depth of concern need not register as strength of occurrent desire, vehemence of asserted opinion, or degree of inferential articulation in one's account of the subject of one's concern—in short, that the texture of an agent's first-personal experience of herself, her history, her actions, and her circumstances is *not* a reliable guide to her rationality or her virtue and, as such, a poor compass by which to steer moral assessment of her actions.

Part of what makes it possible to find outright counterexamples surfacing in her suggestions that moral agency is more complex than we might have thought is the depth of the moral knowledge she displays, almost casually, in her work with the examples that ground her project. Moreover, the motley character of those examples strengthens her work considerably. High literary figures like Madame Bovary share the pages with undergraduate Nietzsche enthusiasts who cannot quite manage to practice what they preach. Self-loathing lesbians and contented anorectics rise up against philosophers who put a premium on reflectively-endorsed desires, self-control, identification, and the varieties of autonomy made of combining these things. Garden-variety unconscious motivation and wishful thinking, together with such phenomena as consumer tendency to make purchases based on the shelf position of goods, elude the grasp of standard philosophical accounts of self-deception (or else are made to seem more rare than we know them to be). Entrenched racism and sexism are unwittingly overridden by egalitarian moral perception. Calm heads make bad decisions while overwrought people yielding to impulses arising under conditions of sleep-deprived anguish hit the nail on the head. Arpaly's examples are the kinds of things that *anyone* capable of grasping the psychological dimensions of a popular novel or movie, of following a talk show discussion of human experience, or, for that matter, of tracking the psychological lives of her intimates, can recognize. That such examples can be drawn from anywhere and everywhere is amply demonstrated by the fact that Arpaly's cases have no one kind of source. As Arpaly puts it, "My cases...display the sort of murkiness and complexity which is *the rule*, not the exception, in life, and especially in the part of life interesting to us as moral philosophers" (30). If the threat posed by Arpaly's insistence that we take such material to heart prompts a defensive "Why care about such hard cases?" the answer she offers is: because these cases are ubiquitous to our subject matter. And her skill in presenting and discussing them ought to carry the day.

Her skill will *have* to carry the day, actually. For although she is clearly right to urge that meeting the biggest challenges posed by duly complex cases will require that we be scrupulous about distinguishing problems in practice from philosophical problems, the relevant form of theory-practice distinction is at once unfamiliar and difficult. Arpaly can accept going philosophical theories of rationality and morality, provided that these resist unrealistic assumptions about first-person authority. But the difficulty of divorcing ourselves from our faith in first-person authority may well be difficulty enough to discourage us from taking up the tasks that she sets for us. The motivation to do engage her project comes from a conviction that the hard work of doing so is worthwhile. That conviction, in turn, builds over the course of the book through Arpaly's treatment of examples.

Like any interesting new project, Arpaly's revisits some familiar questions in new guises. She holds that people deserve praise for acts of good will, blame for acts of ill will, or acts that arise from the absence of good will, and that the degree of praise or blame warranted varies with the depth of motivation or the degree of moral indifference. She also holds that good willing is noninstrumentally attached to performing right acts for right reasons. She glosses this as the claim that the good will as such seeks "to perform actions that have whatever property it is that makes actions right" (115) but leaves open the questions *what* property that might be, and how to account for acting for reasons *de re* more generally. Some of us may doubt that there is any one sort of right-making *property* borne by all and only morally desirable actions. There seems to be no reason to suppose that a singular action-property-based account of moral desirability can be made to work. It is not clear why Arpaly thinks that any such account *could* be in the offing, given the complexity of moral life. There are various other points where one might quibble. One could favor a more traditional sense of "principle" than the sort at issue in her title and urge that virtue is highly principled on her account (the account has everything to do with the rational source of morally desirable action, after all). Then too, given that Arpaly bases her account of moral worth on the quality of an agent's will, coining the term "inverse *akrasia*" to describe an agent who does the right thing for the right reasons against her considered judgment is misleading. The inverse *akratic* will sound; it is the agent's *thinking* or *conscience* that is misguided. These are minor complaints.

A serious question, however, haunts Arpaly's work. At her best, she sets things up in such a way that the point under discussion is perfectly obvious—Who could *fail* to see that an anorectic regime of self-control meets various explicit, formal philosophical criteria for autonomous action, betrays powerful capacities for self-governance, is too variously entrenched to be the result of a single false belief (or a small cluster of false beliefs), coheres powerfully with reflectively endorsed anorexic self-image, and is, for all that, irrational? Who among us is *ever* in a position to answer, definitively and authoritatively, a question about whether she is merely being neurotic (or is instead giving serious matters all due attention) when she is stewing over her career, or her marriage, or her children? Who could *fail* to see that the kind of emotional charge and cognitive complexity *properly* carried by ethically charged circumstances are not the sorts of things that conduce to confidence in one's own authority over one's actions or character? On the one hand, the reader's confidence in moral common sense is bolstered by Arpaly's exploration of its contours. On the other, he cannot *but* be struck by how this very common sense seems to have gone missing in moral philosophy. We will want an account of moral common sense *given* that we will *not* be able to build up the account piecemeal (starting from individual, avowed conviction), *given* that accurate moral perception is sometimes at odds with custom and upbringing, *given* that reflective endorsement, consent, and the like do not provide the resources necessary to explain moral common sense—in short, *given* that most of the initially attractive ways of thinking about moral epistemology will fall prey to *exactly* the kinds of objections that Arpaly levels against standard philosophical work in moral psychology. Arpaly leaves it to us to answer this question.

Still, her book will repay close study for anyone interested in moral psychology, practical reason, virtue, moral motivation, or rational agency. Happily, *Unprincipled Virtue* is such an engaging read that it actively *recruits* the close study that it merits.

CANDACE VOGLER

University of Chicago

NEW BOOKS: ANTHOLOGIES

BENJAMIN, ANDREW, ed.: *Walter Benjamin and Art.* New York: Continuum, 2005. vii, 300 p. Cloth $105.00, paper $24.95.

BERNECKER, SVEN, ed.: *Reading Epistemology: Selected Texts with Interactive Commentary.* Malden, MA: Blackwell, 2006. x, 198 p. Paper $29.95.

BLESSING, KIMBERLY A., and PAUL J. TUDICO, eds.: *Movies and the Meaning of Life.* Chicago: Open Court, 2005. xvi, 302 p. Paper $17.95.

BOROS, JÀNOS, ed.: *Mind in World: Essays on John McDowell's Mind and World (with McDowell's Replies).* Pécs: Brambauer, 2005. 205 p. Paper.

BRAUN, R., and D. KRIEGER, eds.: *Einstein—Peace Now! Visions and Ideas.* Hoboken, NJ: John Wiley, 2005. 313 p. Cloth $27.95.

BRENDER, NATALIE, and LARRY KRASNOFF, eds.: *New Essays on the History of Autonomy: A Collection Honoring J.B. Schneewind.* New York: Cambridge, 2004. ix, 214 p. Cloth $70.00.

BROADIE, ALEXANDER, ed.: *The Cambridge Companion to the Scottish Enlightenment.* New York: Cambridge, 2003. xvi, 366 p. Cloth $65.00.

BROCK, GILLIAN, and HARRY BRIGHOUSE, eds.: *The Political Philosophy of Cosmopolitanism.* New York: Cambridge, 2005. x, 262 p. Cloth $70.00, paper $24.99.

BRONSON, ERIC, ed.: *Poker and Philosophy.* Chicago: Open Court, 2006. xvi, 226 p. Paper $17.95.

BROOK, ANDREW, and KATHLEEN AKINS, eds.: *Cognition and the Brain: The Philosophy and Neuroscience Movement.* New York: Cambridge, 2005. x, 430 p. Cloth $90.00, paper $26.99.

BUCHANAN, IAN, and GREGG LAMBERT, eds.: *Deleuze and Space.* Toronto: University Press, 2005. vii, 245 p. Cloth $89.00, paper $29.95.

BUCKNER, S. CLARK, and MATTHEW STATLER, eds.: *Styles of Piety: Practicing Philosophy after the Death of God.* New York: Fordham, 2006. xi, 303 p. Cloth $75.00, paper $26.00.

BUGH, GLENN R., ed.: *The Cambridge Companion to the Hellenistic World.* New York: Cambridge, 2006. xxix, 371 p. Cloth $80.00, paper $29.99.

BURLEY, JUSTINE, ed.: *Dworkin and His Critics.* Malden, MA: Blackwell, 2004. xxiii, 412 p. Paper $29.95.

BYRON, MICHAEL, ed.: *Satisfying and Maximizing: Moral Theorists on Practical Reason.* New York: Cambridge, 2004. ix, 245 p. Cloth $70.00, paper $26.99.

CALARCO, MATTHEW, and PETER ATTERTON, eds.: *Animal Philosophy: Essential Readings in Continental Thought.* New York: Continuum, 2004. xxv, 221 p. Paper $27.95.

CAMPBELL, JOSEPH KEIM, MICHAEL O'ROURKE, and DAVID SHIER, eds.: *Freedom and Determination.* Cambridge: MIT, 2004. vi, 329 p. Paper $35.00; *Law and Social Justice.* Cambridge: MIT, 2005. x, 355 p. Cloth $70.00, paper $35.00.

CAPPS, JOHN M., and DONALD CAPPS, eds.: *James and Dewey on Belief and Experience.* Urbana: Illinois UP, 2005. x, 290 p. Cloth $45.00, paper $20.00.

CAPUTO, JOHN D., and MICHAEL J. SCANLON, eds.: *Augustine and Postmodernism: Confessions and Circumfession.* Bloomington: Indiana UP, 2005. viii, 264 p. Cloth $60.00, paper $24.95.

CARLSON, ALLEN, and ARNOLD BERLEANT, eds.: *The Aesthetics of Natural Environments.* Orchard Park, NY: Broadview, 2004. 312 p. Paper $22.95.

CARMAN, TAYLOR, and MARK B. N. HANSEN, eds.: *The Cambridge Companion to Merleau-Ponty.* New York: Cambridge, 2005. ix, 396 p. Cloth $75.00, paper $27.99.

CARROLL, THOMAS C., JR., ed.: *Social Ethics: Classical and Applied.* Dubuque, IA: Kendall/Hunt, 2006. xx, 668 p. Paper $120.00.

CARRUTHERS, PETER, STEPHEN LAURENCE, and STEPHEN STICH, eds.: *The Innate Mind: Structure and Contents.* New York: Oxford, 2005. 429 p. Cloth $74.00, paper $35.00.

CHAPMAN, SIOBHAN, and CHRISTOPHER ROUTLEDGE, eds.: *Key Thinkers in Linguistics and the Philosophy of Language.* New York: Oxford, 2005. xii, 282 p. Cloth $65.00, paper $24.95.

CHRISTMAN, JOHN, and JOEL ANDERSON, eds.: *Autonomy and the Challenges to Liberalism.* New York: Cambridge, 2005. xii, 383 p. Cloth $85.00.

CLARK, CHRIS, ed.: *Ways of Knowing: Science and Mysticism Today.* Charlottesville, VA: Imprint Academic, 2005. vii, 260 p. Paper $34.90.

CLAYTON, PHILIP, and PAUL DAVIES, eds.: *The Re-Emergence of Emergence.* New York: Oxford, 2006. xiv, 329 p. Cloth $99.00.

COHEN, ANDREW I., and CHRISTOPHER HEATH WELLMAN, eds.: *Contemporary Debates in Ethics.* Malden, MA: Blackwell, 2005. xii, 348 p. Cloth $69.95, paper $34.95.

COHEN, MARTIN, ed.: *The Essentials of Philosophy and Ethics.* New York: Hodder Arnold, 2006. xv, 310 p. Paper $19.95.

COPP, DAVID, ed.: *The Oxford Handbook of Ethical Theory.* New York: Oxford, 2006. xiv, 665 p. Cloth $74.00.

CRAIG, EDWARD, ed.: *The Shorter Routledge Encyclopedia of Philosophy.* New York: Routledge, 2005. xxvi, 1077 p. Cloth $45.00.

CRARY, ALICE, and SANFORD SHIEH, eds.: *Reading Cavell.* New York: Routledge, 2006. x, 262 p. Cloth $105.00, paper $31.95.

CUDD, ANNE E., and ROBIN O. ANDREASEN, eds.: *Feminist Theory: A Philosophical Anthology.* Malden, MA: Blackwell, 2005. x, 432 p. Cloth $74.95, paper $34.95.

DANCY, R. M., ed.: *Plato's Introduction of Forms.* New York: Cambridge, 2004. xii, 348 p. Cloth $75.00.

DARBY, DERRICK, and TOMMIE SHELBY, eds.: *Hip Hop and Philosophy: Rhyme2Reason.* Chicago: Open Court, 2005. xvii, 221 p. Paper $17.95.

DASTON, LORRAINE, and GREGG MITMAN, eds.: *Thinking with Animals: New Perspectives on Anthropomorphism.* New York: Columbia, 2005. vii, 230 p. Cloth $49.50, paper $25.00.

DAVIES, BRIAN, and BRIAN LEFTOW, eds.: *The Cambridge Companion to Anselm.* New York: Cambridge, 2004. xiii, 323 p. Cloth $65.00, paper $29.99.

DAVIS, CRESTON, JOHN MILBANK, and SLAVOJ ZIZEK, eds.: *Theology and the Political: The New Debate.* Durham, NC: Duke, 2005. xii, 476 p. Cloth $99.95, paper $29.95.

DAVIS, STEVEN, and BRENDAN S. GILLON, eds.: *Semantics: A Reader.* New York: Oxford, 2004. xiv, 921 p. Cloth $85.00, paper $45.00.

DE HAAS, FRANS, and JAAP MANSFELD, eds.: *Aristotle's 'On Generation' and 'Corruption I'.* New York: Oxford, 2004. ix, 347 p. Cloth $74.00.

DE PAULO, CRAIG J. N., PATRICK MESSINA, and MARC STIER, eds.: *Ambiguity in the Western Mind.* New York: Peter Lang, 2005. xv, 248 p. Paper $32.95.

DECKER, KEVIN S., and JASON T. EBERL, eds.: *Star Wars and Philosophy.* Chicago: Open Court, 2005. xlv, 215 p. Paper $17.95.

DENKER, ALFRED, and MICHAEL VATER, eds.: *Hegel's Phenomenology of Spirit: New Critical Essays.* New York: Humanity, 2003. 359 p. Paper $35.00.

DEVITT, MICHAEL, and RICHARD HANLEY, eds.: *The Blackwell Guide to the Philosophy of Language.* Malden, MA: Blackwell, 2006. x, 446 p. Cloth $89.95, paper $39.95.

DIMOVA-COOKSON, MARIA, and W. J. MANDER, eds.: *T.H. Green: Ethics, Metaphysics, and Political Philosophy.* New York: Oxford, 2006. viii, 321 p. Cloth $74.00.

DOBSON, ANDREW, and ROBYN ECKERSLEY, eds.: *Political Theory and the Ecological Challenge.* New York: Cambridge, 2006. vii, 262 p. Cloth $75.00, paper $29.99.

DOLE, ANDREW, and ANDREW CHIGNELL, eds.: *God and the Ethics of Belief: New Essays in Philosophy of Religion.* New York: Cambridge, 2005. xii, 282 p. Cloth $60.00.

DREIER, JAMES, ed.: *Contemporary Debates in Moral Theory.* Malden, MA: Blackwell, 2006. xxiv, 331 p. Paper $34.95.

EILAN, NAOMI, CHRISTOPHER HOERL, TERESA MCCORMACK, and JOHANNES ROESSLER, eds.: *Joint Attention: Communication and Other Minds.* New York: Oxford, 2005. xii, 332 p. Cloth $99.00, paper $35.00.

ELLIOTT, CARL, and TOD CHAMBERS, eds.: *Prozac as a Way of Life.* Chapel Hill: North Carolina UP, 2004. 213 p. Cloth $39.95, paper $19.95.

ERNELING, CHRISTINA E., and DAVID MARTEL JOHNSON, eds.: *The Mind as a Scientific Object.* New York: Oxford, 2005. xiii, 549 p. Cloth $98.50.

ERTESCHICK-SHIR, NOMI, and TOVA RAPOPORT, eds.: *The Syntax of Aspect: Deriving Thematic and Aspectual Interpretation.* New York: Oxford, 2005. xx, 309 p. Cloth $120.00, paper $45.00.

FEENBERG, ANDREW, and DARIN BARNEY, eds.: *Community in the Digital Age: Philosophy and Practice.* Lanham, MD: Rowman and Littlefield, 2004. vii, 295 p. Cloth $79.00, paper $34.95.

FERREIRÓS, JOSE, and JEREMY J. GRAY, eds.: *The Architecture of Modern Mathematics: Essays in History and Philosophy.* New York: Oxford, 2006. xii, 442 p. Cloth $69.50.

FIRESTONE, CHRIS L., and STEPHEN R. PALMQUIST, eds.: *Kant and the New Philosophy of Religion.* Bloomington: Indiana UP, 2006. xxvi, 270 p. Cloth $65.00, paper $24.95.

FRASCA-SPADA, MARINA, and P. J. E. KAIL, eds.: *Impressions of Hume.* New York: Oxford, 2005. 308 p. Cloth $74.00.

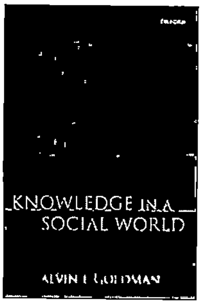
FRIEDMAN, MARILYN, ed.: *Women and Citizenship.* New York: Oxford, 2005. xi, 220 p. Cloth $74.00, paper $24.95.

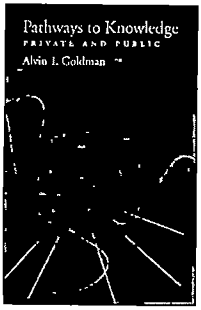
FRIEDMAN, MICHAEL, and ALFRED NORDMANN, eds.: *The Kantian Legacy in Nineteenth-Century Science.* Cambridge: MIT, 2006. vi, 370 p. Cloth $45.00.

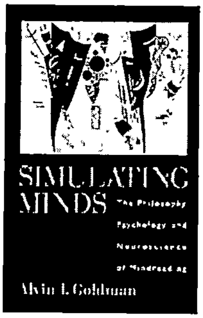
FOLTZ, BRUCE V., and ROBERT FRODEMAN, eds.: *Rethinking Nature: Essays in Environmental Philosophy.* Bloomington: Indiana UP, 2004. viii, 357 p. Cloth $60.00, paper $24.95.

FUGLSANG, MARTIN, and BENT MEIER SØRENSEN, eds.: *Deleuze and the Social.* Edinburgh, UK: University Press, 2006. vi, 280 p. Cloth $90.00, paper $28.00.

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Knowledge in a Social World

ALVIN I. GOLDMAN

"Alvin Goldman, the premier epistemologist of the past two decades, has written a pioneering book that will define the field of social epistemology. Scholars will learn from his judicious and lucid proposals, and they will be wrestling for years with the exciting and important problems he raises."—Philip Kitcher

1999 paper $37.50 cloth $168.00

Pathways to Knowledge

Private and Public

ALVIN I. GOLDMAN

"Alvin Goldman...has made as large a contribution (in both senses) to the theory of knowledge, in all its varieties, as any living philosopher, and to its social aspect in particular in his *Knowledge in a Social World* (1999) and in part III of his *Pathways to Knowledge* (2002). He is the most substantial of analytic social epistemologists."—Anthony Quinton, *Episteme: A Journal of Social Epistemology*

Alvin Goldman examines public and private methods or "pathways" to knowledge, arguing for the epistemic legitimacy of private and introspective methods of gaining knowledge, yet acknowledging the equal importance of social and public mechanisms in the quest for truth.

2002 (paper 2004) $22.00

Simulating Minds

The Philosophy, Psychology, and Neuroscience of Mindreading

ALVIN I. GOLDMAN

"This impressive book is sufficiently clear and accessible to serve as an introduction to simulation theory yet is full of new ideas. It is very clearly the work of a philosopher yet provides an extraordinarily informative guide to the relevant literature in cognitive science and neuroscience. Indeed, for everyone who is interested in our ability to understand each other as minded creatures, *Simulating Minds* is essential reading."—Martin Davies, University of Oxford

(Philosophy of Mind Series)

2006 $35.00

STUDIES IN ANALYTIC PHILOSOPHY
Series Editor: **Quentin Smith, Western Michigan University**

REFERENCE AND ESSENCE
Second Edition
Nathan U. Salmon
433 pp • ISBN 1-59102-215-0 • PB • $35

"Reference and Essence *is one of the true classics in philosophy of language and metaphysics in the past 25 years. It is a must read for any serious student of direct reference and the logic and metaphysics of modality in the post-Kripkean tradition."* **—Scott Soames**
Professor of Philosophy, University of Southern California

THE ONTOLOGY OF TIME
L. Nathan Oaklander
366 pp • ISBN 1-59102-197-9 • PB • $35

"Oaklander's detailed responses to his many allies and opponents down the years also make his book an indispensable guide to the recent history of work in the philosophy of time."
—D. H. Mellor, Emeritus Professor of Philosophy, Cambridge University

THE DOUBLE CONTENT OF ART
John Dilworth
350 pp • ISBN 1-59102-235-5 • HC • $59

"Entirely original, a landmark achievement."
—Dominic McIver Lopes, University of British Columbia

FORTHCOMING

GOD AND METAPHYSICS
Richard M. Gale
350 pp • ISBN 1-59102-269-X • HC • $59

This collection of seminal articles by Richard M. Gale represents the areas of philosophy to which he has made significant contributions—God, time, nonbeing, and pragmatism.

PHILOSOPHY OF RELIGION, PHYSICS, AND PSYCHOLOGY
Essays in Honor of Adolf Grünbaum
Edited by Aleksandar Jokic
300 pp • ISBN 1-59102-369-6 • HC • $59

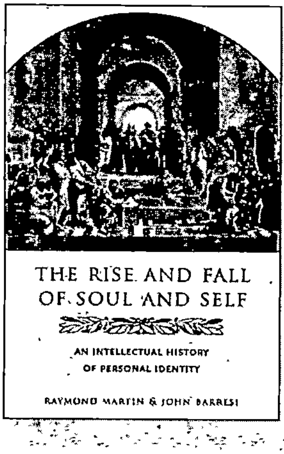
This work contains essays by Nancy Cartwright, Adolf Grünbaum, Quentin Smith, Michael Tooley, Bas van Fraassen, and others, on the areas in philosophy of religion and philosophy of science that Adolf Grünbaum's recent work has significantly influenced.

THE KALAM COSMOLOGICAL ARGUMENT FOR GOD
Mark R. Nowacki
330 pp • ISBN 1-59102-473-0 • HC • $59

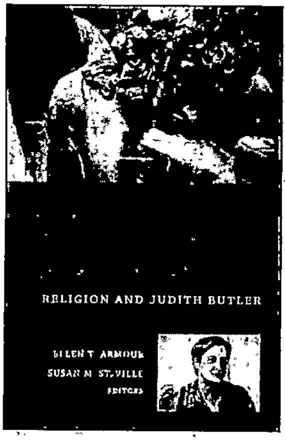
This book breaks new ground by arguing that future progress on the Kalam Cosmological Argument (KCA) requires that the argument be situated within a substance-based metaphysics and a modal theory sensitive to the possibilities and necessities that obtain for substances. With these conceptual tools, thought-experiments supporting the KCA can be properly assessed.

Prometheus Books
59 John Glenn Drive / Amherst, NY 14228
Toll Free: 800-421-0351 / Fax: 716-691-0137 / www.prometheusbooks.com

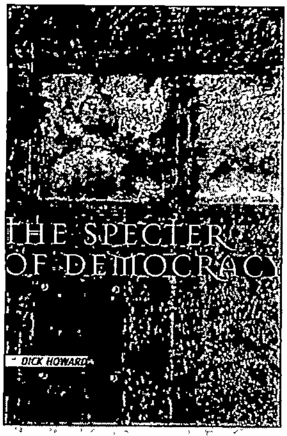
COLUMBIA, *Refer to us*
Read book excerpts at www.columbia.edu/cu/cup


THE RISE AND FALL OF SOUL AND SELF
AN INTELLECTUAL HISTORY OF PERSONAL IDENTITY
RAYMOND MARTIN & JOHN BARRESI

The Rise and Fall of Soul and Self
An Intellectual History of Personal Identity
Raymond Martin & John Barresi
From Plato and Aristotle to Freud and Foucault, this new history traces the efforts of philosophers and psychologists to understand the soul and the self.
cloth - $29.50



Bodily Citations
Religion and Judith Butler
Ellen T. Armour & Susan M. St. Ville
Gender, Theory, and Religion Series
"Lively and incisively intelligent ... a lucid account of the key notions in Judith Butler's thought — among them, subjectivity, performativity, abjection, agency, resistance, subversion, and the materializing power of the body."
— Janet Gyatso, *Harvard University*
paper - $24.50



NOW IN PAPER
The Specter of Democracy
What Marx and Marxists Haven't Understood and Why
Dick Howard
"A shrewd new philosophy for the political left."
— Paul Berman, author of *A Tale of Two Utopias*
paper - $22.50

CAMBRIDGE

BEST IN SCHOLARSHIP

Christ and Horrors
The Coherence of Christology
Marilyn McCord Adams
Current Issues in Theology
$85.00: Hardback: 0-521-86682-0: 344 pp.
$29.99: Paperback: 0-521-68600-8

Nietzsche: 'On the Genealogy of Morality' and Other Writings
Second Edition
Edited by Keith Ansell-Pearson
Translated by Carol Diethe and
Friedrich Nietzsche
Cambridge Texts in the History of Political Thought
$50.00: Hardback: 0-521-87123-9: 216 pp.
$16.99: Paperback: 0-521-69163-X

Kant's 'Critique of Pure Reason'
An Introduction
Jill Vance Buroker
Cambridge Introductions to Key Philosophical Texts
$75.00: Hardback: 0-521-85315-X: 336 pp.
$24.99: Paperback: 0-521-61825-8

Political Theory and the Ecological Challenge
Edited by Andrew Dobson and
Robyn Eckersley
$75.00: Hardback: 0-521-83810-X: 270 pp.
$29.99: Paperback: 0-521-54698-2

Modal Logic for Philosophers
James W. Garson
$80.00. Hardback: 0-521-86367-8: 472 pp.
$34.99: Paperback: 0-521-68229-0

Aquinas, Aristotle, and the Promise of the Common Good
Mary Keys
$70.00: Hardback: 0-521-86473-9: 272 pp.

Prices subject to change.

Principles of Constitutional Design
Donald S. Lutz
$80.00: Hardback: 0-521-86168-3: 278 pp.

Arguing about Gods
Graham Oppy
$90.00: Hardback: 0-521-86386-4: 472 pp.

Gandhi's Philosophy and the Quest for Harmony
Anthony J. Parel
$75.00: Hardback: 0-521-86715-0: 240 pp.

Socratic Virtue
Making the Best of the Neither-Good-Nor-Bad
Naomi Reshotko
$85.00: Hardback: 0-521-84618-8: 218 pp.

The Cambridge Companion to Early Modern Philosophy
Edited by Donald Rutherford
Cambridge Companions to Philosophy
$75.00: Hardback: 0-521-82242-4: 438 pp.
$29.99: Paperback: 0-521-52962-X

Human Goodness
Pragmatic Variations on Platonic Themes
Paul Schollmeier
$80.00: Hardback 0-521-86384-8: 320 pp.

The Heart of Judgment
Practical Wisdom, Neuroscience, and Narrative
Leslie Paul Thiele
$80.00: Hardback: 0-521-86444-5: 334 pp.

Moral Repair
Reconstructing Moral Relations after Wrongdoing
Margaret Urban Walker
$70.00: Hardback: 0-521-81088-4: 264 pp.
$27.99: Paperback: 0-521-00925-1

www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

FROM CAMBRIDGE

Fundamentals of Critical Argumentation
Douglas Walton
Critical Reasoning and Argumentation
$75.00: Hardback: 0-521-82319-6
$29.99: Paperback: 0-521-53020-2: 360 pp.

The Cambridge Companion to Tocqueville
Edited by Cheryl B. Welch
Cambridge Companions to Philosophy
$80.00. Hardback: 0-521-84064-3
$29.99: Paperback: 0-521-54996-5: 464 pp.

CAMBRIDGE COMPANIONS ONLINE

The Online Collection: A Major New Resource

Cambridge Companions Online is a major new online resource for the humanities. The online collection enhances and complements the book series by offering students unrivalled access and functionality.

- **The Complete Companions Collection**
  covering literature, classics, philosophy, religion and cultural studies and currently offering more than 220 titles
- **The Literature and Classics Companions Collection**
  currently offering more than 135 titles
- **The Philosophy, Religion and Culture Companions Collection**
  currently offering more than 85 titles

*Sign up for your **FREE INSTITUTIONAL TRIAL** today by writing*
ccol@cambridge.org

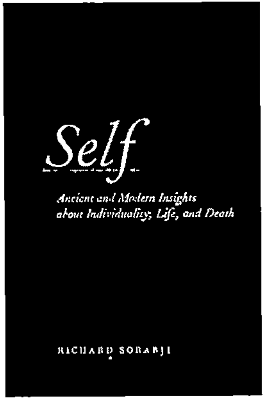
Prices subject to change

www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

PHILOSOPHY

NEW FROM RICHARD SORABJI


Self
Ancient and Modern Insights about Individuality, Life, and Death
RICHARD SORABJI

Self

Ancient and Modern Insights
about Individuality, Life, and Death

Richard Sorabji

"Richard Sorabji's books typically display a remarkable combination of virtues: meticulous scholarship, amazing historical range, philosophical insight and precision, and a vivid sense of the issues that a nonphilosophical reader will find interesting and engaging. *Self* may be his best, displaying all those virtues at a very high level."—Martha Nussbaum

Cloth $35.00

NOW IN PAPERBACK

Aristotle on Memory

Second Edition

"Anyone concerned with Aristotle's psychology, theory of mind, or rhetoric, anyone interested in mnemonic systems, and anyone trying to work out for himself a theory of memory, should read Aristotle's treatise *On Memory*, with the comments by Richard Sorabji."—Anthony Preus, *International Studies in Philosophy*

Paper $15.00

Necessity, Cause and Blame

Perspectives on Aristotle's Theory

"It is difficult to convey the extraordinary richness of this book. . . . A Greekless philosopher could read it with pleasure. . . . At the same time, its learning and scholarship are enormous."—G. E. M. Anscombe, *Times Literary Supplement*

Paper $19.00

Time, Creation and the Continuum

Theories in Antiquity and the Early Middle Ages

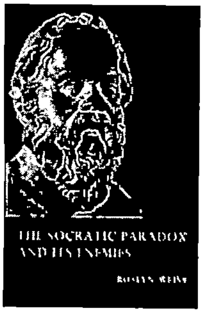
"Splendid. . . . The canvas is vast, the picture animated, the painter nonpareil. . . . Sorabji's work will encourage more adventurers to follow him to this fascinating new-found land."—Jonathan Barnes, *Times Literary Supplement*

Paper $22.50

The University of Chicago Press • www.press.uchicago.edu

PHILOSOPHY

NEW FROM CHICAGO



THE SOCRATIC PARADOX AND ITS ENEMIES

Roslyn Weiss

"I expect this book to stir controversy of the best sort: the kind that acknowledges a challenge worth responding to. This is an eye-opening, indeed quite startling reinterpretation of Socratic claims that have for some decades been the prize exhibits for the standard view of Socrates. Reading along, I often felt like cheering."—G. R. F. Ferrari, University of California, Berkeley

CLOTH $35.00

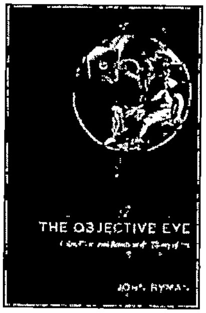
CONFRONTING ARISTOTLE'S ETHICS

Ancient and Modern Morality

Eugene Garver

"This is a smart, inventive, highly original, and important contribution to our understanding of Aristotle. Eugene Garver has a clear and unified vision of Aristotle's project in ethics and, in particular, of the importance of Aristotelian politics and metaphysics in this understanding. His is an entirely new and refreshing approach, one that challenges many of the mainstream accounts of Aristotle."—Charles Young, Claremont Graduate School

CLOTH $49.00



THE OBJECTIVE EYE

Color, Form, and Reality
in the Theory of Art

John Hyman

"John Hyman combines formidable skills as an analytical philosopher with extensive knowledge of the art historical tradition to reach genuine insights into the powers and limits of pictorial representation. The book tackles fundamental questions about the very possibility of visual art and never fails to illuminate the multiple topics it addresses. I am sure it will rapidly establish itself as a canonical work in the field."—Peter Lamarque, editor of the *British Journal of Aesthetics*

8 COLOR PLATES, 97 HALFTONES
PAPER $35.00

NOW IN PAPERBACK

MEMORY, HISTORY, FORGETTING

Paul Ricoeur

Translated by Kathleen Blamey and David Pellauer

"Moving. . . . It provides strong ethical advice for today as official commemoration brings the memory of some European genocides to the fore while maintaining a persistent silence on others."—Luisa Passerini, *Times Higher Education Supplement*

PAPER $25.00

THE UNIVERSITY OF CHICAGO PRESS • www.press.uchicago.edu

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details.

University
Batta
14·5·07

THE JOURNAL OF PHILOSOPHY

VOLUME CIII, NUMBER 10
OCTOBER 2006

page
481 *Intrasubjective Intentional Identity* — Walter Edelberg
503 *What Externalists Should Say about Dry Earth* — Daniel Z. Korman
521 *Disjunctive Causes* — Carolina Sartorio

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Libraries and Institutions (print and online)	$180.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.

POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2006 by the Journal of Philosophy, Inc.

ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CIII, NO. 10, OCTOBER 2006

INTRASUBJECTIVE INTENTIONAL IDENTITY*

Some forty years ago, Peter Geach posed what he called the problem of *intentional identity*:[1]

(1) Hob thinks a witch blighted Bob's mare, and Nob thinks she killed Cob's sow.

One reading of this sentence, the *Geach* reading as I shall call it, can be true even if there are no witches, and even if neither Hob nor Nob has mistaken any particular person for a witch. So understood, the sentence seems to say that Hob's and Nob's attitudes are in some sense *about the same witch*, even though she does not (necessarily) exist. This reading, Geach argued, lies beyond the scope of conventional intensional formal languages for propositional attitudes.[2] If Geach is right about that, and if such languages mark the limit of what we understand about the logic of attitude ascriptions, then there is something about their logic that escapes us. Further, if Geach is right, and if such languages can be used to represent any proposition that your preferred semantical framework deems intelligible, then there is something about the content of propositional attitudes that lies beyond the scope of your preferred semantical views.

---

*For helpful comments and discussion, I wish to thank Lisa Downing, Sam Fleischacker, Dorothy Grover, Bill Hart, David Hilbert, Nick Huggett, Peter Hylton, Mitzi Lee, Abe Roth, Marya Schechtman, and an anonymous referee for this JOURNAL.

[1] Geach, "Intentional Identity," this JOURNAL, LXIV, 20 (October 26, 1967): 627–32. For simplicity I have replaced Geach's 'Nob wonders whether...' with 'Nob thinks that...'.

[2] By a conventional intensional formal language for the attitudes, I mean (roughly) one (a) that contains—in addition to the usual individual constants, predicates, truth-functional connectives, and apparatus for quantification and identity—operators or predicates for belief, desire, and other attitudes; and (b) for which all quantifier expressions carry existential import. I say "roughly" to allow for certain variations that do not affect the argument here. Often I will refer to these as conventional (or standard) intensional languages. Classic versions in a possible worlds semantical framework are analogous to those provided for alethic modality in Saul Kripke, "Semantical

0022-362X/06/0310/481–502

I shall review Geach's argument in the next section. My primary aim in this paper is to argue that Geach's problem is considerably more interesting—and more disturbing—than we realized. I argue that three strategies for deflecting the philosophical impact of Geach's puzzle—here called *Skepticism*, *Reductivism*, and *Reasoned Indifference*—are woefully inadequate. A fourth strategy posits nonexistent entities (Hob's and Nob's witch, for instance) as new semantical determinants. Yet this strategy, which I here call (*Intentional*) *Realism*, is generally regarded as something to be avoided at all costs. So we are in a nasty predicament.

Let me begin with an overview of the four strategies. The *Realist* strategy is the idea that the truth conditions of the Geach reading of (1) are best expressed by means of sentences of the form of (2) or (3).[3]

(2) $\exists\alpha$ [Hob believes (α is a witch & α blighted Bob's mare) & Nob believes (α killed Cob's sow)]

(3) $\exists\alpha\exists\beta$ [Hob believes (α is a witch & α blighted Bob's mare) & Nob believes (β killed Cob's sow) & $\alpha \approx \beta$]

Here, the quantifiers '$\exists\alpha$' and '$\exists\beta$' are interpreted referentially (rather than substitutionally), but they lack existential import. That is, they range over entities that in an appropriate sense "do not really exist." In some versions of this approach, the quantifiers in (2) and (3) are said to range over *intentional objects*, and for convenience I will adopt that term here, although I wish to be as noncommittal as possible about the nature of the relevant quantifier domains. The analysis given by (2) posits intentional objects that are global in the sense that they can "appear" in the attitudes of more than one person. The analysis given by (3) posits intentional objects that are local in the sense that they can appear in the attitudes of only

---

Considerations on Modal Logic," *Acta Philosophica Fennica*, XVI (1963): 83–94. For early work on this sort of approach, see Jaakko Hintikka, *Knowledge and Belief* (Ithaca: Cornell, 1962). For possible worlds versions with an applicative (function and argument) semantics, see David Lewis, "General Semantics," *Synthèse*, XXII (1970): 18–67; and Richard Montague, "The Proper Treatment of Quantification in Ordinary English," in Hintikka, Michael Moravcsik, and Patrick Suppes, eds., *Approaches to Natural Language* (Boston: Reidel, 1973), pp. 221–42. For versions that utilize logically structured propositions, see Lewis, "General Semantics," section v; and Scott Soames, "Direct Reference, Propositional Attitudes, and Semantic Content," *Philosophical Topics*, XV (1987): 47–87. There are important syntactic and semantic differences among such languages, yet none that will affect the arguments to be advanced here.

[3] Geach's sentence poses at least two distinct, but related, questions: (a) What is the formal syntax of the sentence? (b) What sort of formal language is required simply in order to state the truth conditions of the sentence? Sentences (2) and (3) are offered as an answer to the second question, not as an answer to the first. The suggestion that (2) or (3) correctly state the truth conditions for (1) would seem compatible with a number of theories as to its formal syntax.

one person, and posits a counterpart relation, $\approx$, among the local intentional objects.

For a variety of reasons, most contemporary philosophers find the Realist approach exceedingly unpalatable. It can be argued (a) that it violates a useful criterion of ontological commitment; (b) that it is ontologically extravagant; (c) that by countenancing nonexistent objects, it faces well-known logical difficulties; (d) that we lack a decent metaphysical theory of the postulated intentional objects; (e) that we lack principles of existence (or subsistence), identity, and predication for the postulated intentional objects; (f) that by postulating a domain of intentional objects mediating the relation between mind and world, the approach takes us too close to the problematic epistemology of Descartes and Locke; and (g) that we lack a theory of the intentional relation expressed implicitly by the double in-quantification in (2), and explicitly by the predicate '$\approx$' in (3). (Notice that (g) means that this approach cries out for a theory of mind-to-mind intentional relations, in addition the various mind-to-world intentional relations we usually countenance.) I hope I have not left anything out.

The three deflective strategies aim to skirt all of those difficulties. *Reductivists* argue that with some ingenuity we *can* after all express the truth conditions for the Geach reading of (1) in conventional intensional formal languages.[4] The idea is that a proper theory of the pronoun in the Geach reading of (1) will reveal it to be synonymous (at least relative to a context of use) with another English sentence in which the pronoun has been replaced by an appropriate definite, indefinite, or universal noun phrase. The latter sentence is to be *well-behaved* in the sense that its truth conditions can be expressed within the framework of standard intensional languages. *Skeptics* have argued in various ways that sentence (1) simply lacks the reading Geach claims for it.[5] The *Indifferent* do not care. This reaction can be perfectly reasonable, so it is possible to speak of *Reasoned Indifference.* When you have a well-supported theory, recalcitrant data are sometimes best simply ignored, especially if they are of little or no independent interest. No one has ever argued that Geach's sentence is central to our philosophical, scientific, or other important concerns. It is hard to see what significant role the sentence could play in anything we really care about. So if the Skeptical and Reductivist strategies prove to be indefensible, and with a host of philosophical problems and puzzles facing the Realist strategy, the most reasonable course is surely to express passing interest in Geach's sentence, and then simply to *ignore* it as a bit of anomalous data.

---

[4] See, for instance, Daniel C. Dennett, "Geach on Intentional Identity," this JOURNAL, LXV, 11 (May 30, 1968): 335–41; and Montague, "The Proper Treatment of Quantification in Ordinary English." For an interesting and extremely well-developed Reductivist approach, see Jeffrey C. King, "Intentional Identity Generalized," *Journal of Philosophical Logic,* XXII (1993): 61–93, and "Anaphora and Operators," *Philosophical Perspectives,* VIII, *Logic and Language* (1994): 221–50.

[5] Dennett, "Geach on Intentional Identity"; and Montague, "The Proper Treatment of Quantification in Ordinary English" utilize a combined strategy. Reductive analyses are provided for certain versions of the Geach reading; skeptical arguments are then advanced against putative readings that do not succumb to the reduction.

My main aim in this paper is to try to sharpen and intensify the conflict between the Realist strategy and its deflective alternatives. Earlier I cataloged the most pressing problems for the Realist approach. I want to balance the scales by explaining what is deeply unsatisfying with the reactions of Skepticism, Reductivism, and Indifference. At the most salient level, I will be occupied with Indifference. I will argue that certain intentional identity sentences play crucially important roles in common-sense psychological explanations of human behavior. I shall also argue that these intentional identity statements are *essential* in performing these roles. That is, other, well-behaved attitude ascriptions cannot take over these functions. If successful, these arguments will ultimately show that all three deflective strategies are highly problematic. My hope is that the arguments will not only provoke renewed defenses of Skepticism, Reductivism, and Indifference, but stimulate a more serious consideration of Realism and how it could be provided with adequate philosophical foundations.[6]

I. OVERVIEW OF GEACH'S PUZZLE

Readings of Geach's sentence (1) can be distinguished by entailment patterns. On one reading, Geach's sentence entails both that a real witch exists, and that Hob and Nob have beliefs about the same real object. This is the so-called *de re* reading of the sentence, and it can be represented in conventional intensional formal languages by a sentence like (4):

(4) $\exists x[x$ is a witch & Hob believes (x blighted Bob's mare) & Nob believes (x killed Cob's sow)]

---

[6] It is not entirely clear that problems (a) through (g) are insuperable. On problem (a) see, for instance, William P. Alston, "Ontological Commitments," *Philosophical Studies*, IX (1958): 8–17; and Susan Haack, "Quantifiers," in *Philosophy of Logics* (New York: Cambridge, 1978), pp. 39–55. On problems (c) and (d), see Terence Parsons, *Nonexistent Objects* (New Haven: Yale, 1980). On problems (c), (d), (e), and (g), see my "Intentional Identity and the Attitudes," *Linguistics and Philosophy*, XV (1992): 561–96, and "A Perspectivalist Semantics for the Attitudes," *Noûs*, XXIX, 3 (1995): 316–42.

I mention the *de re* reading of (1) only to set it aside. I shall also set aside a neighboring *de re* construction carrying only the second of the two entailments I just mentioned:

(5) $\exists x$[Hob believes (x is a witch & x blighted Bob's mare) & Nob believes (x killed Cob's sow)]

An altogether different reading of (1) carries neither of the two entailments. This reading can be true in situations like the following:

Story 1. Bob's mare is sick, and Hob and Nob were called in for a joint diagnosis. "A witch blighted her," Hob says. "Yes," Nob replies, "and I'll bet she killed Cob's sow." In reality, Bob's mare has a congenital defect that is beginning to take its toll.

Story 2. Yesterday's newspaper carried a story about a witch named Helga who is on something of a rampage. Hob and Nob don't know one another at all, but each reads the newspaper story. Hob thought to himself, "I'll bet Helga blighted Bob's mare." Nob thought, "I'm sure Helga killed Cob's sow." But Helga doesn't really exist.

I shall call this the *Geach reading* of the sentence, or the *Geach sentence* for short. (One might also call it the *de dicto* reading, but that term would prejudge certain theoretical issues. The most we can say is that Geach's reading is *not* the *de re* one.)

Notice that the first conjunct of the Geach sentence falls easily within the scope of conventional semantical theories. For sentence (6) below can be analyzed as (7), often called the *de dicto* analysis of (6).

(6) Hob believes a witch blighted Bob's mare.
(7) Hob believes $\exists x$(x is a witch & x blighted Bob's mare)

This is important. That Hob has a belief about something that does not exist causes no problem at all for conventional theories. It is rather the intersubjective import of Geach's sentence that causes the problem for the standard accounts. To see this, notice first that we would normally formalize (8) in first order logic as in (9).

(8) A witch blighted Bob's mare and she killed Cob's sow.
(9) $\exists x$(x is a witch & x blighted Bob's mare & x killed Cob's sow)

That is, we would construe the pronoun 'she' as functioning like a variable, bound by the quantifier phrase 'a witch'.[7] The problem is that there is no way to combine the techniques of analysis we have applied to (6) and to (8) to arrive at a plausible analysis of the Geach sentence. For instance, (10) will not do:

---

[7] The bound-variable analysis of (8) is even more apparent in Montague-style grammars for English. See, for instance, Montague, "The Proper Treatment of Quantification in Ordinary English."

(10) Hob believes $\exists x[x$ is a witch & x blighted Bob's mare & Nob believes (x killed Cob's sow)]

The last occurrence of 'x' in (10) is bound by the quantifier, but to achieve this we have placed 'Nob believes' in the scope of 'Hob believes'. This analysis of the Geach sentence does not entail that Nob believes anything at all, so it is clearly unacceptable. An additional problem with (10) is that it entails that Hob has beliefs about Nob and about what Nob believes. As story 2 shows, the Geach sentence can be true even if Hob has no beliefs at all about Nob. (Sentences (4) and (5) above likewise construe the pronoun in (1) quantificationally, but these are *de re* constructions and do not capture the Geach reading.)

Alternative proposals abandon the idea that the pronoun in the Geach sentence is quantificational. One of these supposes that it is as a *pronoun of laziness*: a pronoun that goes proxy for a definite description constructible from its antecedent and the surrounding discourse. On this proposal one could take the Geach sentence to have the same truth conditions as (11), where the definite description takes narrowest possible scope:

(11) Hob believes a witch blighted Bob's mare, and Nob believes that *the witch that Hob believes blighted Bob's mare* killed Cob's sow.

The intended reading of (11) can be expressed in standard intensional languages as (12):

(12) Hob believes $\exists x(x$ is a witch & x blighted Bob's mare) & Nob believes $\exists x$[Hob believes (x is a witch & x blighted Bob's mare) & $\forall y$(Hob believes (y is a witch & y blighted Bob's mare) $\supset y = x$) & x killed Cob's sow]

This, of course, is an attempt to implement the Reductivist strategy. It should be clear that this particular proposal fails. Sentence (12) is true only if Nob has beliefs about Hob and what Hob believes; yet in story 2 the Geach reading of (1) is true even though neither Hob nor Nob knows of the other's existence.[8] Other Reductivist hypotheses can be advanced, and as the discussion unfolds I will be discussing some of these—implicitly for most of the paper, but explicitly at the end. The prospects for Reductivism, we shall see, are rather dim.

---

[8] On pronouns of laziness, see Peter Geach, *Reference and Generality* (Ithaca: Cornell, 1962), and "Referring Expressions Again," *Analysis*, XXIV (1964): 73–92. In "Intentional Identity," Geach considers and rejects the laziness analysis of (1), largely for the reasons stated above.

If the truth conditions of the Geach sentence cannot be represented in conventional intensional formal languages for the attitudes, no proposition expressible in these languages would be what the Geach sentence asserts. Given a few plausible assumptions, it would follow that when the Geach sentence is true, the propositions believed by Hob or Nob, or both, are not expressible in these languages.[9] In that case, conceptions of propositions normally thought to provide adequate foundations for theories of content for the attitudes would have been shown to be insufficient for that purpose.

Let us take stock. Unless the Reductivist strategy can be successfully implemented, the truth conditions of the Geach sentence—and arguably the propositions Hob or Nob, or both, believe when it is true—cannot be expressed in languages that otherwise seem as powerful as one could wish for. Problems (a) through (g) eagerly await us if we try to extend the language and the underlying theories of content in the way the Realist suggests. No wonder, then, that people have reacted with skepticism about the linguistic intuitions that generate the problem. Or that some have argued that the Geach sentence ultimately does reduce, by principles governing the pronoun, to a sentence that falls within the scope of standard intensional formal languages and associated theories of propositional content. Or that the sentence generates only passing interest, thereafter to be largely ignored.

II. INTRASUBJECTIVE CASES

When he introduced the puzzle, Geach provided no hint as to what roles intentional identity statements might play in natural language: what purposes they might serve, what utility they might have. He simply argues that we cannot account for our linguistic intuitions about the sentence. Pursuing the problem in this fashion, it quickly becomes hard to see it as anything more than a logical curiosity—and one safely ignored. Some later illustrations of intentional identity, such as this example from Mark Richard, seem to fit that view of the significance of the problem:[10]

---

[9] Call a proposition expressible in conventional intensional formal languages for the attitudes an *ordinary* proposition, and a proposition that is not ordinary in this sense an *extraordinary* one. Then, as a first approximation, the following assumptions would suffice to establish, from the fact that the Geach sentence expresses an extraordinary proposition, that Hob or Nob, or both, believes an extraordinary proposition when the Geach sentence is true: (i) that the Geach sentence contains two belief ascriptions with Hob and Nob, respectively, as subjects; (ii) that belief ascriptions express relations between persons and propositions; and (iii) that the syntactic rules used to generate the Geach sentence cannot generate sentences expressing extraordinary propositions from sentences expressing ordinary ones.

(13) Hob thinks the unicorn ate the petunias, but Nob thinks it probably didn't.

Considered in isolation from other kinds of cases, examples like Geach's (1) and Richard's (13) lend considerably plausibility to the strategy of Reasoned Indifference. Yet these illustrations create a false impression of the importance and difficulty of the problem. Intentional identity comes in both intersubjective and intrasubjective forms, and the latter play important roles in explaining human behavior, while still posing all the problems of philosophical analysis we find in Geach's intersubjective example. Take the case of Grandma, for instance.

> *Story 3.* Grandma is stomping off toward the barn with her rifle. She looks very upset. "What's wrong, Grandma?" you ask. "It's a snake. A dad-blasted snake and it's eating the chickens. I'm going to hunt it down and shoot it." But there is no snake. The missing chickens have run off to play with the chickens on a neighboring farm.

We ask: Why is Grandma headed toward the barn, gun in hand? Within the framework of commonsense psychology, (14) provides a perfectly good explanation:

(14) Grandma thinks there's a snake in the barn, and she wants to shoot it.

The reading we want here is not the *de re* reading given by (15), nor the neighboring *de re* construction (16), since Grandma's belief is clearly not about any particular snake, nor even about any particular real object to the effect that it is a snake in the barn.

(15) $\exists x$[x is a snake & Grandma believes (x is in the barn) & Grandma wants (Grandma shoots x)]

(16) $\exists x$[Grandma believes (x is a snake & x is in the barn) & Grandma wants (Grandma shoots x)]

Nor will sentence (17) do, since it does not entail that Grandma wants anything, but only that she *believes* she does.

(17) Grandma believes $\exists x$[x is a snake & x is in the barn & Grandma wants (Grandma shoots x)]

---

[10] Richard, *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them* (New York: Cambridge, 1990), pp. 4–5.

On the Realist approach, we would state the truth conditions of the Grandma sentence (14) along the lines of (18) or (19):

(18) $\exists\alpha$[Grandma believes (α is a snake & α is in the barn) & Grandma wants (Grandma shoots α)]
(19) $\exists\alpha\exists\beta$[Grandma believes ($\alpha$ is a snake & α is in the barn) & Grandma wants (Grandma shoots β) & $\alpha \approx \beta$]

As in the intersubjective application of the Realist approach, these quantifiers range only over intentional objects that, in an appropriate sense, do not exist. (When the Realist approach is applied as in (19), we assume that the intentional objects are local not only in the sense that they cannot appear in the attitudes of more than one person, but also in the sense that they cannot appear in more than one attitude-type of a single person.) All the earlier objections (a) through (g) to the Realist approach apply here—except that (g) now becomes: (g′) we lack a theory of the intrasubjective intentional relation expressed implicitly by the double in-quantification in (18), and explicitly by the predicate '≈' in (19). So we reach for the three deflective strategies here as much as in the case of Hob and Nob.

Unlike Geach's (1) and Richard's (13), the Grandma sentence (14) is actually useful in explaining human behavior. Such intrasubjective intentional identity statements are perfectly natural explanations of human behavior in terms of belief and desire. (Indeed, similar intrasubjective intentional identity statements might be useful in explaining (other) *animal* behavior. You hold your closed hand out to your dog, the way you always do when you have a treat for him. Why is he pawing at your hand and trying to open it? "He thinks you have a treat in your hand, and he wants to eat it." Only you are teasing him; your hand is empty.) The Skeptical response—that sentences (1) and (13) lack the reading Geach claimed for them—seems utterly preposterous in the case of (14). The reaction of Reasoned Indifference—that we need not provide a philosophical theory of the problematic readings of (1) or (13), since they serve no clear explanatory or other purpose—seems completely implausible in the case of (14).

Yet perhaps we can explain Grandma's behavior by means of some *well-behaved* sentence. The most promising idea would be to try to explain her actions by means of a sentence in which the pronoun in (14) has been replaced by a descriptive noun phrase (NP) of some sort. I call this the *descriptivist* explanatory strategy, and I shall consider three kinds of descriptivist candidate explanations: those in which the pronoun has been replaced by (i) an indefinite NP (*a* K), (ii) a definite NP (*the* K), or (iii) a universal NP (*every* K), where K is any common noun phrase (such as 'snake' or 'snake such that it is in the barn'). In each case, the common noun phrase K may be either (a) *plain* (that is, free of doxastic expressions such as 'believes' or 'thinks'), or (b) *doxasticized* (containing at least one doxastic expression). Thus 'snake in the barn' is plain, 'snake believed by Grandma to be in the barn' is doxasticized. I extend this terminology to cover NPs formed by combining common noun phrases K with various determiners such as 'a', 'the', and 'every'. Combining (i), (ii), and (iii) with (a) and (b), we have six kinds of descriptivist candidates to consider all together.

We begin by first considering descriptivist candidates that rely on *plain* NPs:

(20) Grandma believes there's a snake in the barn. Grandma wants to shoot *a snake in the barn.*
(21) Grandma believes there's a snake in the barn. Grandma wants to shoot *the snake in the barn.*
(22) Grandma believes there's a snake in the barn. Grandma wants to shoot *every snake in the barn.*

For short I shall refer to these as the *plain indefinite, definite,* and *universal* candidates, respectively. I shall postpone discussion of the universal candidate (22) for a later section, and focus for now on (20) and (21).

One problem with explanations like (20) and (21) is that they are subject to what I shall call the *Distribution Problem.* On the desired reading, the second sentences in (20) and (21) are expressed in standard intensional languages as (20a) and (21a), respectively:

(20a) Grandma wants $\exists x(x$ is a snake & x is in the barn & Grandma shoots $x)$
(21a) Grandma wants $\exists x[x$ is a snake & x is in the barn & $\forall y(y$ is a snake & y is in the barn $\supset y = x)$ & Grandma shoots $x]$

(We need the *de dicto* readings here, since the *de re* readings of (20) and (21) are false in story 3.) The trouble is that (20a) and (21a) are also false in the story. For assuming that desire distributes over conjunction—a principle I defend in the next section—(20a) and (21a) both entail (23):

(23) Grandma wants $\exists x(x$ is a snake & x is in the barn)

Yet (23) is false in the story. For (23) says that Grandma wants it to be the case that there is at least one snake in the barn. Grandma despises most snakes. She certainly does not want any in the barn. Since the intended readings of (20) and (21) entail (23), and (23) is false in story 3, it follows that (20) and (21) are false in story 3. So (20) and (21) cannot explain Grandma's behavior in that story. (More precisely, the argument is this: given the Distribution Principle defended in the next section, and given that the intended readings of (20) and (21) fall in the scope of conventional intensional languages, these two sentences each entail (23), which is clearly false in story 3. So either the Distribution Principle is false, or the intended readings of (20) and (21) lie outside the scope of conventional intensional languages, or (20) and (21) are false in story 3 and cannot explain Grandma's behavior.[11])

We now turn to candidate explanations that appeal to *doxasticized* NPs. For short I shall refer to these as *doxasticized* indefinite, definite, and universal candidates:

(24) Grandma thinks a snake is in the barn. Grandma wants to shoot *a snake such that she thinks it is in the barn.*
(25) Grandma thinks a snake is in the barn. Grandma wants to shoot *the snake such that she thinks it is in the barn.*
(26) Grandma thinks a snake is in the barn. Grandma wants to shoot *every snake such that she thinks it is in the barn.*

Again, I want to postpone discussion of the universal candidate (26) and to focus for now on (24) and (25). These are subject to a doxasticized version of the Distribution Problem. For the intended (narrow scope) readings of the second sentences in (24) and (25) both entail (27):

(27) Grandma wants $\exists x[x$ is a snake & Grandma believes (x is in the barn) & Grandma shoots $x]$

Assuming again that desire distributes over conjunction, (27) entails (28):

(28) Grandma wants $\exists x[x$ is a snake & Grandma believes (x is in the barn)$]$

Yet (28) is false. Grandma wants no such thing. Yes, she believes a snake is in the barn. But she does not really *want* there to be a snake such that she believes it to be in the barn. "There are some things you'd just as soon not know about, and this snake he's one of them. My heart, you know. But since he's there and I've got my chickens to think about, well I just have to go on and shoot it." In some versions of the story, there will be a different reason why (28) is false. Suppose Grandma knows that when it comes to snakes in the barn, she is very keen and rarely issues false positives. She knows that if she believes there is a snake in the barn, this is very likely true. She believes there is a snake in the barn, but strongly prefers the situation in which the barn is snake-free and in which she accordingly does not believe there is a snake in the barn (and she strongly prefers that situation to the one in which the barn is snake-free but in which she falsely believes there to be a snake). So she does not want it to be the case that she believes there is a snake in the barn. It seems that no definite or indefinite candidate explanation, whether plain or doxasticized, can successfully explain Grandma's behavior in story 3.

---

[11] I would suggest that (20) is simply false in the story. On the other hand, (21) seems ambiguous between a false reading given by (21a), and a true reading in which the definite description 'the snake in the barn' is wholly anaphoric on the indefinite description 'a snake'. On the latter reading, (21) is little more than a stylistic variant of (14), and like (14) it lies beyond the scope of conventional semantical theories.

III. THE DISTRIBUTION PRINCIPLE

The argument I used to generate the Distribution Problem relied on the assumption that desire distributes over conjunction:

(29) S wants $(p \,\&\, q)$ $/ \therefore$ S wants p

At first there seem to be counterexamples. One might suggest, for instance, that (30) does not entail (31), and similar examples are easy to come by.

(30) I want to put on a parachute and jump out of a plane.
(31) I want to jump out of a plane.

The desire expressed in (30) is conditional—I want to jump out of a plane, *but only if I (first) put on a parachute*—and (31) neglects to mention this, so (31) is false.

Or so one might argue. This argument against the Distribution Principle relies on the assumption that where the content of desire is conditional, it is false to ascribe a desire with that content unless you explicitly mention the condition. Since it is not clear how we should understand the logical form of conditional desire ascriptions,[12] we would do better to formulate this assumption by saying that argument form (32) is valid:

(32) I want $(p \,\&\, q)$
$\sim$ [I want $(\sim p \,\&\, q)$]
$\therefore$ $\sim$ (I want q)

---

[12] '*I want A, but only if B*'. Does this have the form $WA \supset B$? Or $W(A \supset B)$? Should that material conditional be a counterfactual conditional? Or is this kind of conditional desire to be formalized with a binary conditional desire operator, as $W(A//B)$? How does tense figure in? '*If B, then I want A*'. Does this have the form $B \supset WA$? Or $W(B \supset A)$? Should that material conditional be a counterfactual conditional? Or is this kind of conditional desire to be formalized with a binary conditional desire operator, as $W(A/B)$? At least for the latter sort of conditional desire report, the problems and theoretical options seem analogous to those for conditional obligation, on which there is an extensive literature: for an overview and references, see Risto Hilpinen, "Deontic Logic," in Lou Goble, ed., *The Blackwell Guide to Philosophical Logic* (Malden, MA: Blackwell, 2001), pp. 159–82; see especially §8.5.

If argument form (32) is valid, one could then argue that since (30) and (33) below are both true, (34) is also true; hence (31) above is false.

(30) I want to put on a parachute and jump out of a plane.
(33) I do not want to not put on a parachute and jump out of a plane.
(34) I do not want to jump out of a plane.

The trouble with such arguments against the Distribution Principle is that they are easy to stand on their head. I want to put on a parachute and jump out of a plane—yes, but only if the ripcord is attached to the plane, and only if the plane does not lurch suddenly just as I am about to jump, and only if I do not land in a tree, and only if I am not hit on the way down by another airplane, and.... The problem is that if argument form (32) is valid, then (30) itself is going to be false—and it is going to be very difficult indeed to find a finite replacement.

The more general point, of course, is that most of our desires are conditional. I want to go rock climbing, but not if I break my neck in the process. You want to buy a new camera, but not if you will lose it the very next day. He really wants to eat this hamburger, but not if he will choke to death on it. In each case, the conditional desire can be expressed in terms of a conjunction: He wants to eat this hamburger, but *he does not want to eat this hamburger and choke on it.* It would be ludicrous to infer from the italicized conjunct that he does not want to eat the hamburger. The unconditionalized statement is true—he wants to eat this hamburger—and similarly for the other cases. So, I want to suggest, if you *really do* want to put on a parachute and jump out of a plane, then you do want to jump out of a plane. The Distribution Principle is valid.[13]

This is one of those many places where it is important to distinguish between information pragmatically imparted by a use of a sentence in a context, and the proposition semantically expressed by the sentence in the context.[14] Background information and context can affect how we revise our understanding of a person's psychological states when we accept a new report of her desires. The therapist hears his depressed patient say, "I want to jump out of a plane. I want to overdose on pills." The skydiving instructor, after a presentation to recruit new students, hears someone say, "I want to jump out of a plane; I want to feel the exhilaration." The therapist comes to believe the patient desires to jump out of a plane without a parachute; the skydiving instructor comes to believe the prospective student wants to jump out of a plane while wearing a parachute. Yet in both cases, the unadorned report of desire is perfectly true. The content of the unadorned report, together with background information and context, can provide the audience with information that goes beyond the content of the report itself.

---

[13] The Conjunction Principle is different: S wants p, S wants q / $\therefore S$ wants $(p \,\&\, q)$. This inference is clearly invalid. You can want your brother to come to Thanksgiving dinner and want your sister to come to Thanksgiving dinner without wanting them both to attend.

[14] See, for instance, Paul Grice, "Logic and Conversation," in Donald Davidson and Gilbert Harman, eds., *The Logic of Grammar* (Encino, CA: Dickenson: 1975), pp. 64–75; Saul Kripke, "Speaker's Reference and Semantic Reference," in Theodore Uehling, Peter French, and Howard Wettstein, eds., *Midwest Studies in Philosophy,* Volume 2, *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: Minnesota UP, 1977), pp. 64–75, especially §3; Jon Barwise and John Perry, *Situations and Attitudes* (Cambridge: MIT 1983), especially pp. 258–64; Nathan Salmon, *Frege's Puzzle* (Cambridge: MIT, 1986); Soames, "Direct Reference, Propositional Attitudes, and Semantic Content," especially §v, and "Substitutivity" in Judith Jarvis Thomson, ed., *On Being and Saying: Essays in Honor of Richard Cartwright* (Cambridge: MIT, 1987), pp. 99–132.

Another reason for accepting the Distribution Principle looks to the way desires are processed in instrumental reasoning, and is much more speculative. Suppose that this processing proceeds by subgoaling.[15] In particular, suppose that if a person desires that $(p \,\&\, q)$, the processing of that desire automatically generates subgoals p and q. If subgoals—at least those introduced in this specific way from conjunctive desires—are themselves desires, then the Distribution Principle must hold.

IV. UNIVERSAL NOUN PHRASE CANDIDATES

I claimed earlier that intentional identity statement (14) provides a good, common-sense explanation of Grandma's behavior in story 3, and that no well-behaved attitude ascription provides as good an explanation.

(14) Grandma thinks there's a snake in the barn, and she wants to shoot it.

In section II, I argued that proposed explanations that appeal to narrow scope definite or indefinite NPs, whether plain or doxasticized, fail because they fall prey to the Distribution Problem. I postponed discussion of proposed explanations that replace the recalcitrant pronoun with a narrow scope *universal* NP, such as plain (22) or doxasticized (26).

(22) Grandma thinks a snake is in the barn. Grandma wants to shoot *every snake in the barn.*

(26) Grandma thinks a snake is in the barn. Grandma wants to shoot *every snake such that she believes it to be in the barn.*

---

[15] See Elaine Rich, *Artificial Intelligence* (New York: McGraw-Hill, 1983), pp. 25–53; and Keith J. Holyoak, "Problem Solving," in Daniel N. Osherson and Edward E. Smith, eds., *Thinking* (Cambridge: MIT, 1990), pp. 117–46.

On the readings we want, the truth conditions of (22) and (26) are represented by (22a), and (26a), respectively:

(22a) Grandma believes $\exists x$(x is a snake & x is in the barn). Grandma wants $\forall x$[(x is a snake & x is in the barn) $\supset$ Grandma shoots x].

(26a) Grandma believes $\exists x$(x is a snake & x is in the barn). Grandma wants $\forall$x[Grandma believes that (x is a snake & x is in the barn) $\supset$ Grandma shoots x)]

The Distribution Problem is obviously no difficulty whatsoever for (22a) and (26a). So universal NP candidates like (22) and (26) would seem to allow us to explain Grandma's behavior without intentional identity statements like (14). For those who wish to pursue the strategy of Reasoned Indifference, this greatly increases the interest and importance of the universal NP candidates. Yet (22) and (26) are themselves subject to other serious objections. (Some of these objections will apply with equal force against the definite and indefinite NP candidates considered earlier.)[16]

The Specificity Problem. This is best explained by an illustrative example:

> *Story 4.* Grandma has a pet garter snake whom she calls "Jake." All too often, Jake slips out of his cage and slithers off to the barn to eat bugs. She thinks Jake is probably in his cage, but she acknowledges that there's also a fair chance he's out in the barn. If Jake is in the barn, she certainly doesn't want to shoot him. Grandma mistakenly believes there is a wild snake in the barn, and that it's after the chickens. She wants to shoot it.

Here, intentional identity statement (14) still provides a good explanation of Grandma's behavior. Yet (22) is false: it is simply not the case that Grandma wants to shoot every snake in the barn, on the reading given by (22a). A slight variant of story 4 will show that the doxasticized version (26) will not do, either. Suppose Grandma *does* believe that Jake is in the barn, but she certainly does not want to shoot *him.* Intentional identity statement (14) will still explain Grandma's behavior in the variant version, but (26) will be false on the reading given by (26a).[17]

---

[16] Notice that (22) is equivalent to "Grandma believes there's a snake in the barn. Grandma wants it to be the case that if there is a snake in the barn, then she shoots it." So the arguments against the universal NP candidate (22) apply equally well against the conditionalized version. Similar remarks apply to (26) and its conditionalized version.

[17] Notice that "I dropped a coin and it fell into the gutter" is true even if you dropped two coins, and only one of them fell into the gutter. Similarly, sentence (14) is true even if Grandma thinks there are two snakes in the barn, and she wants to shoot only one of them.

Of course we might be able to supplement the descriptive content in the relevant NPs, to home in on the right snake in story 4 and its variant:

(35) Grandma thinks a *wild* snake is in the barn. Grandma wants to shoot every object such that it is (or she believes it to be) a *wild* snake in the barn.

Yet in an extended version of story 4, in which Grandma likes wild garter snakes and thinks that there could (also) be one of these in the barn, we would have to complicate the explanation further:

(36) Grandma thinks a snake *such that it is wild but not a garter* is in the barn. Grandma wants to shoot *every snake in the barn such that it is wild but not a garter.*

If we try to explain Grandma's behavior along these lines, we must include in our explanation each such qualification, as Grandma's beliefs and desires about snakes become more nuanced. Some might be sanguine about the prospects for solving the Specificity Problem in this way. But there is good reason to be skeptical. More often than we would like to think, our attitudes toward objects are determined by features of our mental life that we are insufficiently aware of for them to enter into the contents of our beliefs and desires: our moods, emotions, the degree of stress we are under, whether we are well-rested or tired, our general level of happiness, background noise or other distractions, associations inaccessible to consciousness, and so on. Suppose you are meeting someone for the first time. Your attitudes toward that person, including some persistent desires, can be determined to a large extent by these sorts of factors. In that case, you might have persisting desires about that person, even though these factors—for instance, that at the time of meeting you were in a very good (or dreadful) mood that you were calm and at peace (or totally stressed out), or that the person unconsciously reminded you of a childhood classmate you loved (or hated)—do not enter into the content of any of your beliefs and desires. Love and hate probably work more in this way than we would ever like to admit.

Suppose that this is how it is with Grandma and her snake: the one she wants to shoot. What distinguishes the snake she wants to shoot from other snakes is simply that at the time she came to believe it existed, she was in a certain mental state. Suppose she is so unaware of being in that state that none of her attitudes has as any part of its content that she was in that state at that time. Here, it will not do to supplement the universal noun phrase in (22) or (26) with predicates like 'wild', or 'chicken-eating' or anything else of this sort.

One might try to solve the Specificity Problem by introducing into the universal NP a complex predicate that singles out the right snake by referring to the episode in which the belief was first acquired, perhaps by referring to the time of that episode:

(37) Grandma believes there is a snake in the barn. There is a time *t* such that Grandma first came to have this belief at time *t*, and Grandma wants it to be the case that she shoots *every object x such that she first came to believe at t that x is a snake in the barn.*

One problem with (37) is that this explanation will work only if Grandma has the concepts of time and belief, since these concepts figure in the content of her desire according to (37). Such an assumption might work in the case of Grandma, but it will not work if we are explaining the behavior of a very small child. A second problem emerges if Grandma is mistaken about the time at which she first acquired the belief in question. Suppose she first acquired a belief that a snake (Jake, for instance) was in the barn at 10:00 a.m., but did not want to shoot that snake. At 11:00 a.m. the same day, she first acquired a belief that there was a "second" snake in the barn, and because of her mood and other such factors, she acquired a persistent desire to shoot that snake. Suppose she gets it backwards: she believes, of 10:00 a.m., that this is when she first acquired the belief about the snake she wants to shoot; and she believes, of 11:00 p.m., that this is when she first acquired the belief about the snake she does not want to shoot. In that case, (14) will still explain why Grandma is headed toward the barn with her rifle, but (37) will be false. A third problem is that (37) seems at best a highly misleading explanation of Grandma's behavior. For it suggests that what is important *to Grandma* about the snake, her reason for wanting to shoot it, is that she first became aware of its existence at a certain time.

Clearly, the Specificity Problem poses similar difficulties for the definite and indefinite candidates we considered earlier.

The Problem of Targeted Action. Another problem with the universal NP replacements emerges when we consider cases like the following one.

> *Story 5.* When Grandma wants to clear the barn of wild snakes, she takes her new rifle because it reloads more quickly and doesn't have such a mean recoil. She does this even if she believes there is just one wild snake in the barn. "You might think there's just one," she says, but you can never tell when you might find more." When on the other hand she wants to kill just one particular wild snake she believes to be there, she takes her grandfather's rifle which for sentimental reasons she enjoys using more than the new one. On the present occasion, Grandma takes her grandfather's rifle. "I'll get him with this for sure," she says. She believes there is just one snake, but it is compatible with her beliefs that there are others.

The problem is that for any given kind *K*, an agent may act quite differently when the agent wishes to do something to *every instance* of kind *K*, and when she wishes do it to *a particular* instance of kind *K* that she believes to exist. This difference can persist even when the agent currently believes there is only one instance of kind *K*, provided that the agent realizes that she may later revise her beliefs to allow that there is more than one *K*. Intentional identity statement (14) will explain Grandma's behavior in story 5. Universal NP candidates like (22) will not:

(22) Grandma thinks a snake is in the barn. Grandma wants to shoot *every snake in the barn.*

This simple illustration takes the value of *K* to be the property of *being a snake in the barn*, but clearly the problem will persist with other (nondoxasticized) values of *K*. (It may be difficult to argue that the Problem of Targeted Behavior applies to candidates in which *K* is doxasticized in a way that refers to Grandma's present beliefs. Yet below we shall see there is an independent argument against the doxasticized versions.)

The Problem of Case-by-Case Strategies. In some situations, the agent might not have adopted any rule to the effect that every instance of kind *K* is to be shot, because she wants to decide on a case-by-case basis. Sometimes it can be hard to foresee every possibility. "If I find a snake that's not a garter, I'll kill it. But not if it has a collar around its neck indicating that it's someone's pet. But then again, if the collar says the snake belongs to Alvin, I'll shoot it. Unless Alvin apologizes. Unless he apologizes with his fingers crossed...." Suppose that Grandma has adopted a policy that she is going to treat the *K* instances on a case-by-case basis. She believes that a *K* snake is in the barn, and she has decided to shoot "this one." In that case, (14) still provides a good explanation. But because she has decided to treat *K* instances on a case by case basis, sentence (38) will be false:

(38) Grandma thinks that a (*K*) snake is in the barn. Grandma wants it to be the case that she shoots every *K* snake in the barn.

(It is tricky to argue that this problem applies to candidates in which *K* is doxasticized in a way that refers to Grandma's present beliefs. Yet below we shall see there is an independent argument against the doxasticized versions.)

The Practical Inference Problem. This objection applies only to the doxasticized versions. In practical reasoning, we sometimes form intentions on the basis of our beliefs and desires. Grandma does this when she forms an intention to shoot the snake on the basis of her belief that there is a snake in the barn, and her desire to kill it. Let us try to home in on the inference pattern that Grandma relies on in making her inference. The logic of practical inference is of course far less clear than we would like it to be, yet schema (39) represents a form of inference that seems correct in respects that interest us here:

(39) *S* believes that there exists at least one *F*.
S desires that it will be the case that every *F* is *G*.
S believes that for any *x* that is *F*, the most effective means of seeing to it that it will be the case that *x* is *G* is to see to it that *S* does *A* to *x*.
∴ *S* intends to see to it that *S* does *A* to some *F*.

No doubt we would need to add certain provisos to allow for overriding desires and various other factors, but this will not affect the argument I wish to make. Moreover, schema (39) is stated in a form that includes the relevant propositional attitude in each step, as we would if we were explaining or predicting *S's* practical inference. In an actual logic of practical inference, it is less clear that these operators would be included in the premises of the inference.[18] This issue also does not affect the present argument.

Let us see what happens when we try to work with (26) as the basis for our explanation of how Grandma reasons in accord with schema (39).

(26) Grandma thinks a snake is in the barn. Grandma wants to shoot *every snake such that she believes it to be in the barn.*

Here we run into a problem. For (40) below is not an instance of schema (39), since we do not have a consistent value for '*F*' throughout.

(40) Grandma believes that there exists at least one *snake in the barn.* Grandma wants it to be the case every *snake such that she believes it to be in the barn* is killed by her. Grandma believes that the most efficient way to see to it that any *snake in the barn* is killed by her is for her to shoot it. Therefore, Grandma intends to shoot some *snake in the barn.*

Rather, we would need (41):

(41) Grandma believes there is at least one *snake such that she believes it to be in the barn.* Grandma wants it to be the case that every *snake such that she believes it to be in the barn* is killed by her. Grandma believes that any *snake such that she believes it to be in the barn* could most effectively be killed by her if she shoots it. Therefore, Grandma intends to shoot some *snake such that she believes it to be in the barn.*

---

[18] See John Broome, "Practical Reasoning," in José Luis Bermúdez and Alan Millar, eds., *Reason and Nature: Essays in the Theory of Rationality,* (New York: Oxford, 2002), pp. 85–112.

In other words, the doxasticization of desire we see in (26) infects the whole of Grandma's practical inference. This is not just implausible, but wildly so. We would have one (plain) form of practical inference for reasoning about *real* snakes, and another (doxasticized) form for reasoning about snakes that are believed to exist, but do not.

Here is a much simpler account of Grandma's practical reasoning:

(42) Grandma believes that there is a snake in the barn, and she wants to kill it. Grandma believes that she could most effectively kill it by shooting it. Therefore, she intends to shoot it.

Obviously the practical inference schema on which (42) relies involves intentional identities. Yet I conjecture that without intentional identity it will be impossible to provide a plausible account of practical inference. All the more reason for thinking we cannot ignore it.

The Tracking Problem. (This objection applies to definite, indefinite, and universal candidates, whether plain or doxasticized.) We want to be able to explain Grandma's behavior as she tracks the snake (which does not really exist). *Now* Grandma is going up into the hayloft with the rifle. Sentence (43) explains this stage of her tracking behavior perfectly well:

(43) Now Grandma thinks a snake (or the snake) is up in the hayloft, and she (still) wants to shoot it.

If we are attempting to explain her behavior by appealing to universal candidates, we would need something like the following:

(44) Now Grandma thinks that a snake is up in the hayloft. Grandma wants to shoot every snake such that it is up in the hayloft.

As Grandma hunts, her beliefs change: she acquires new beliefs and loses some old ones. "It was in the cow's stall. But when I opened the barn door I heard him slither up into the hayloft." To explain Grandma's behavior by means of universal, definite, or indefinite candidates, we must assume that the content of Grandma's desire changes with each slight alteration in her beliefs about the snake. Explaining Grandma's behavior by means of an intentional identity statement like (14) carries no such commitment. This is an important difference. Grandma's beliefs purport to track the snake. To explain her behavior in terms of her beliefs and desires, these mental states must be coordinated around the "figure" of the snake. The descriptivist approaches we have been considering would appear quite plausible if we assumed that this coordination is achieved (only) through the descriptive content of the relevant beliefs and desires: as Grandma's beliefs about the snake shift, so does the content of her desires, in such a way that the desires and beliefs share certain parts of their descriptive content. Yet this is a substantive assumption about mental processing—and one that seems highly implausible. A more reasonable picture assumes that such beliefs and desires are coordinated not by descriptive content, but by certain causal or functional relationships among aspects of the relevant mental representations. (A system of causally linked tags, for instance, would seem much more efficient than content-matching, when content on the side of belief is rapidly changing.)

V. CONCLUSION

In the opening section I distinguished four possible reactions to Geach's puzzle about intentional identity: Skepticism, Reductivism, Indifference, and Realism. Most of my argument in this paper aimed to defend two theses about the general framework of commonsense psychology:

(a) In this framework, we often explain human (and perhaps other animal) behavior by means of intentional identity statements.
(b) Omitting these statements from the framework would significantly diminish its explanatory scope and power.

The examples offered here in support of these two claims illustrate how implausible is the Skeptical claim that intentional identity statements lack the readings that Geach claimed for them. The arguments show also that the strategy of Reasoned Indifference is in serious trouble, since these sentences do play an important role in commonsense psychology, a role that cannot be played by the well-behaved sentences of the framework.

If correct, the arguments also critically undermine the Reductivist strategy. As I mentioned earlier, Reductivist accounts typically take the form of a hypothesis about the nature of the problematic pronoun in intentional identity statements. According to one, considered briefly earlier, the pronoun is a pronoun of laziness that goes proxy for a narrow scope definite description built up out the antecedent and surrounding discourse. A more liberal hypothesis claims that the pronoun functions as variable bound by a "substitutional quantifier" phrase that takes (narrow scope) definite descriptions as substituends. A third hypothesis proposed by Jeffrey King claims that the pronoun is a "context dependent quantifier."[19] According to this theory, the pronoun functions semantically like a quantifier phrase whose characteristics are determined in a rule-governed way from appropriate features of context. These theories all share the thesis that the truth conditions for (tokens of) intentional identity statements can be expressed in terms of well-behaved propositional attitude statements.[20] To the extent that the theories rest on the underlying idea that intentional identity is to be understood not in terms of causal or functional relations but in terms of identity of descriptive content, they can be seen as vestiges of descriptivist theories of reference and intentionality. Yet if no well-behaved propositional attitude statements can, without a significant loss of explanatory power, replace intentional identity statements in explanations of behavior, these Reductivist theories are in deep trouble. (Indeed, many of the candidate substitutes for intentional identity were not even true in the situations under consideration.)

---

[19] See King, "Intentional Identity Generalized," and "Anaphora and Operators."

These are of course plausibility arguments and are not intended to be absolutely conclusive. Yet too much discussion of intentional identity has proceeded without any attempt to consider how the sentences are actually employed. The unfortunate result has been that we sometimes underestimate the robustness of the intuitions supporting the recalcitrant readings, undervalue their theoretical interest, and misjudge the difficulty of providing a satisfying theoretical account. The Realist approach is, as I mentioned, not without serious problems of its own, yet the alternatives—at least those considered here—are not nearly as attractive as one might have supposed.[21]

WALTER EDELBERG

University of Illinois at Chicago

---

[20] I say "in terms of" because on the substitutional-quantificational interpretation of intentional identity, it is not claimed that some finite well-behaved sentence expresses the truth-condition. Rather, the intentional identity statement is claimed to be true if one or more of its infinitely many well-behaved substitution instances is true.

[21] For reasons of space, I have here limited my focus to how intrasubjective intentional identity figures in explanations of human behavior. In a subsequent paper, I plan to take up how intersubjective intentional identity figures in explaining the behavior of collectives.

WHAT EXTERNALISTS SHOULD SAY ABOUT DRY EARTH*

Twin earths are counterfactual near-duplicates of earth. Externalism about mental content has come to be widely accepted on the basis of intuitions about twin earths. But there is one sort of twin earth, "dry earth," that has been invoked in several arguments against externalism. Dry earth seems to its inhabitants (our intrinsic duplicates) just as earth seems to us, that is, it seems to them as though there are rivers and lakes and a clear, odorless liquid flowing from their faucets. But, in fact, this is an illusion; there is no such liquid anywhere on the planet.[1] There are a number of related objections to externalism concerning the nature of the concept that is expressed by the word 'water' in the mouths of the inhabitants of dry earth. I intend to answer two of them. The first, raised by Gabriel Segal, concerns the application conditions of this concept. The second, raised by Paul Boghossian, concerns the complexity of the concept.[2]

Externalism, as it is to be understood in what follows, is the thesis that (E1) for all natural kinds *K*, it is metaphysically impossible that *K* have instances that differ from actual instances of *K* with respect to their basic physical constitution, and (E2) one cannot possess natural-kind concepts or refer to natural kinds without having had causal interaction with instances of the relevant natural kind. A natural kind is to be understood throughout as a kind all of whose actual instances share a distinctive basic physical constitution. This statement of exter-

---

* Thanks to Kent Bach, Derek Ball, Chad Carmichael, Matti Eklund, Richard Fumerton, Sandy Goldberg, John Hawthorne, Dave Liebesman, Clayton Littlejohn, Brad Majors; Marc Moffett, Brian Rabern, Mark Sainsbury, Laura Schroeter, Thomas Seung, Josh Snyder, and Michael Tye. I am especially grateful to Sarah Sawyer for extremely helpful comments on an early draft of the paper and to George Bealer for many illuminating conversations about the epistemological foundations of externalism.

[1] Please ignore various annoying features of this example, for instance, that there cannot truly be an intrinsic duplicate of one of us whose body contains no water.

[2] Boghossian's argument appears in his "What the Externalist Can Know A Priori," *Proceedings of the Aristotelian Society*, XCVII (1997): 161–75. Segal's argument appears in his *A Slim Book about Narrow Content* (Cambridge: MIT, 2000). Here is one further dry-earth argument against externalism, also due to Segal (pp. 43–44). Even by externalist lights, concept possession presumably supervenes globally on the physical features of a world. Yet, as everything in the supervenience base on dry earth is duplicated on earth, it would follow (*pace* externalism) that we, on earth, must possess the very same concepts as the inhabitants of dry earth. Sarah Sawyer answers this objection in her "Sufficient Absences," *Analysis*, LXIII (2003): 202–08.

0022-362X/06/0310/503–20

nalism will be a bit strong for the tastes of many self-proclaimed "externalists," but it will serve our present purposes: since, as we shall see, the arguments from dry earth fail to undermine even this crude version of externalism, they certainly fail to undermine the weaker versions of the thesis that self-proclaimed externalists would be willing to endorse.[3]

I. APPLICATION CONDITIONS AND DEFAULT CONDITIONALS

Before turning to Segal's and Boghossian's arguments, I shall make some general remarks about externalism and about what externalists are and are not antecedently committed to. This is because Segal and Boghossian each maintain that certain responses to their arguments from dry earth—responses that I wish to defend—constitute a departure from externalism. In particular, I shall be interested in what (if anything) externalists are committed to saying about cases in which what was believed to be a natural kind turns out not to be a natural kind at all.

Externalism is a thesis about natural kinds. We have excellent reasons for believing that water is uniformly composed of H_2O and is therefore a natural kind. If water indeed is uniformly composed of H_2O, then externalist thesis (E1) implies that it is metaphysically necessary that water contain H_2O. However, it might turn out that water is not uniformly composed of H_2O.[4] Scientists might turn out to have been radically mistaken. Water might turn out to be uniformly composed of some other chemical substance, XYZ. In that case, (E1) implies that it is metaphysically necessary that water contain XYZ. Water might even turn out not to be a natural kind at all. Water might turn out to be jade-like, having one chemical composition in the northern hemisphere and another in the southern; water might turn out to be wildly disuniform, each body of water on the planet having its own unique chemical composition; or it may even turn out that there never was any water to begin with, that we have been subject to a grand illusion at the hands of aliens or the government or what have you.

---

[3] A more satisfactory statement of externalism would restrict (E2) to atomic natural-kind concepts and would allow interactions with one's linguistic community to suffice (in a certain range of cases) for the acquisition of natural-kind concepts. Many self-proclaimed externalists will want to restrict (E1) as well (or, alternatively, the definition of 'natural kind') for, as it stands, (E1) has the absurd implication that if it turns out that all hats share some particular chemical composition, then it is impossible for there to have been hats made out of radically different material. Finally, some self-proclaimed externalists are interested in advocating only the austere thesis that individualism is false—that mental content does not supervene on the internal states of individuals—without advocating any particular account of where and why this supervenience failure occurs.

[4] As is customary, I reserve the expression 'it might turn out that' for a familiar variety of epistemic possibility, one that is wholly compatible with metaphysical impossibility.

Should it turn out that in fact water is not a natural kind, (E1) implies nothing of interest about water, nor does (E2) for that matter. Externalists might nevertheless be interested in the application conditions of our concept water—that is, the extension of the concept in the actual world and in counterfactual scenarios—on the off chance that water turns out not to be a natural kind. But, since externalism itself renders no verdict on the application conditions of non-natural-kind concepts, how are externalists to settle the issue? I will suggest one way that externalists might go about it, which fits nicely with the standard externalist practice of consulting intuitions about twin earths.

(Actually, it is not crucial, for the purposes of the paper, that one take intuition to be the means by which we settle these issues. Some externalists may believe that it is by some other cognitive faculty or process that we determine the application conditions of the concept water. Ned Block and Robert Stalnaker, for instance, contend that, if it turns out that we are wrong about the actual composition of the water-like samples on earth, then the application conditions of the concept water are to be settled on the basis of a "decision that will be driven by empirical and theoretical considerations."[5] I have my doubts about whether anything other than intuition—least of all the results of a decision-driven process—can serve as *evidence* in favor of a particular specification of the application conditions of the concept water under alternative empirical suppositions. For the purposes of the paper, however, those externalists who do not think that intuition can play the requisite evidential role may substitute their preferred epistemic resources throughout.)

Let us remind ourselves how externalism was established in the first place. Having presupposed that all of the water on earth is uniformly composed of H_2O, most people have the intuition that a superficially water-like substance on a twin earth that contains no H_2O, but rather is uniformly made up of XYZ molecules, is not water. Externalists call upon this intuition as evidence that water necessarily contains H_2O and, furthermore, that the inhabitants of this twin earth—even though they are our duplicates in all intrinsic respects—do not share our concept water, have no beliefs about water, and use a word 'water' that differs in meaning from our word 'water'. Finally, the best explanation for these differences appears to be that the inhabitants of this twin earth have had no causal interaction with water. It is in this way that externalists have determined the application conditions and acquisition conditions of the concept water, given the knowledge that water is uniformly composed of H_2O. Externalist theses (E1) and (E2) may then be obtained by generalizing upon these results.

---

[5] "Conceptual Analysis, Dualism, and the Explanatory Gap," *Philosophical Review*, CVIII (1999): 1–46, at p. 21. Cf. Laura Schroeter, "Against A Priori Reductions," forthcoming.

Externalists may, in just the same way, determine what the application conditions of our concept water are, should it turn out that we are wrong about the actual composition of water. This can be done by suspending the supposition that the samples of water on earth are all H_2O and, instead, consulting one's intuitions about the relevant twin earths on alternative empirical suppositions. Many will already be familiar with this kind of procedure, having at some point humored a two-dimensionalist by "considering a world as actual."[6] To see that water must be composed of XYZ if the water on earth turns out to be uniformly composed of XYZ, one need only consult one's intuitions about a twin earth on which the water-like samples are composed of something other than XYZ on the empirical supposition that the water on earth is all XYZ. On such a supposition, one has the intuition that samples of H_2O on this twin earth are not samples of water. That our intuitions fluctuate as we vary our suppositions concerning actual empirical facts should come as no surprise. Similarly, in order to determine the application conditions of the concept water on the off chance that water is compositionally highly disuniform, one consults one's intuitions about whether the XYZ on a twin earth is water on the supposition that (actually) each ocean, river, and lake on the planet has a distinct chemical composition: the Pacific Ocean is H_2O, the Atlantic is OPQ, the Mississippi River is RST, and so forth. Intuitively, the XYZ on a twin earth would be water if in fact water has a highly disuniform composition.[7] This suggests that anything superficially water-like would count as water given that water is highly disuniform.

---

[6] That said, one need not accept two-dimensionalism in order to accept all of what I have to say in this paper. One need not accept that these facts about the application conditions of the concept water are somehow parts or "aspects" of the concept water or of what is said in discourse about water. Nor need one accept that these facts about the application conditions of the concept water play any interesting role in determining what is conversationally asserted in discourse about water. One might, if one so chooses, represent some of the observations in the text on a two-dimensional grid—but that does not suffice to make this a two-dimensionalist theory.

[7] I have no intention of "intuition-mongering" here or anywhere else in the paper. Some may not share this intuition (though I have found it to be very widely shared) and may believe that the concept water is necessarily vacuous if it turns out that actual water-

These intuitions may be systematized to obtain a series of "default conditionals," which govern the semantics of the word 'water'.[8] For instance,

(i) If water turns out to be compositionally uniform, then 'water' expresses a concept that applies to all and only samples of that compositional kind with respect to all counterfactual situations.

(ii) If water turns out to have a highly disuniform composition, then 'water' expresses a concept that applies to all and only samples of superficially water-like kinds with respect to all counterfactual situations.[9]

One might expand this list by considering other ways in which we might turn out to be wrong about the constitution of water and consulting our intuitions about the relevant twin earths on the relevant empirical suppositions. For instance, one might suppose that water, like jade, divides into two natural kinds; or one might suppose that, despite appearances, there in fact is no water on earth, that it is all an illusion.

Of particular interest for what follows will be this last default conditional for the off chance that (despite appearances) there turns out to be no water on earth. As before, the content of this default conditional is to be identified by consulting our intuitions about the relevant hypothetical twin earth on the relevant empirical supposition. On the supposition that (despite appearances) there is no water on earth, consider a twin earth on which the clear, odorless, potable, ... liquid has some arbitrary chemical composition, XYZ. For purposes of illustration, let us suppose that that our intuition is that

---

like samples are compositionally disuniform. There is no need to convince them otherwise, since my main point in the present section is that no particular answer is either entailed by, or inconsistent with, externalism.

[8] George Bealer discusses default conditionals in his "A Theory of Concepts and Concept Possession," in Enrique Villanueva. ed., *Concepts* (Atascadero, CA: Ridgeview, 1998), pp. 261–301, at pp. 294–95, and his "Modal Epistemology and the Rationalist Renaissance," in Tamar Szabó Gendler and John Hawthorne, eds., *Conceivability and Possibility* (New York: Oxford, 2002), pp. 71–125, at pp. 108–10. David Chalmers and Frank Jackson discuss such conditionals (under the heading "application conditionals") in their "Conceptual Analysis and Reductive Explanation," *Philosophical Review*, cx (2001): 315–60, at p. 325f.

[9] These default conditionals may need to be modified in light of further intuitions concerning, for example, whether H_2O in the form of ice or a black tar would be water on the relevant suppositions. By design, the resultant default conditionals would exactly track our intuitions (indeterminacies allowed) as we tailor the default conditionals to accommodate more and more detailed and complex variations on the familiar twin-earth scenario.

this liquid *is* water.[10] Then we have reason to accept the following further default conditional:

(iii) If in fact there is no water, then 'water' expresses a concept that applies, with respect to all counterfactual situations, to all and only samples with the superficial features that water was believed to have.

Some externalists (either on the basis of their intuitions, or despite their intuitions) may disagree, maintaining that if there is no water, then 'water' expresses a necessarily vacuous concept. I address this concern in the following section.

Is the conjunction of (i), (ii), and (iii) in any way at odds with externalism? If it is, I cannot see how. Externalist thesis (E1) requires that if water is a natural kind, then all counterfactual samples of water must share its chemical composition. Default conditional (i), by itself, ensures that this condition is satisfied; and (E1) has no bearing whatsoever on (ii) and (iii), since these concern only what would be the case if water turns out *not* to be a natural kind. Externalist thesis (E2) requires that if water is a natural kind, then one must interact with water in order to acquire the concept water or to refer to water. This is consistent with (though not entailed by) default conditional (i) and, like (E1), is entirely irrelevant to (ii) and (iii).[11]

Is there any way of filling in the default conditionals that would be at odds with externalism? Yes. If the consequent of default conditional (i) were to say that 'water' expresses a concept that applies to all and only superficially water-like kinds with respect to all counterfactual situations, this would run afoul of (E1). But it is difficult to see how any particular way of filling in the consequents of default conditionals (ii) or (iii) could be in conflict with externalism—or, for that matter, could even be more or less true to the "spirit" of externalism. These default conditionals can only be more or less true to our intuitions about twin earths. Externalism is a thesis about natural kinds and commits one to a substantive thesis about water in particular only if water turns out to be a natural kind.

---

[10] This, for what it is worth, is my intuition about the case. Even if it turns out that there is no water, surely there at least could have been. For suppose that there is no water on earth; that our hallucination were to wane; and that, years later, a liquid with all of the superficial features that we took water to have were to start raining down on earth. Would it be water? Would the word 'water' apply to this rain when we exclaim: "For the first time in history, there is real water in my tub"? Surely it would.

[11] Cf. Brian McLaughlin and Michael Tye, "Externalism, Twin Earth, and Self-Knowledge," in Barry C. Smith and Cynthia Macdonald, eds., *Knowing Our Own Minds: Essays on Self-Knowledge* (New York: Oxford, 1998), pp. 285–320, at p. 305.

It may be helpful to think of these default conditionals as semantic (or perhaps "metasemantic") principles that reflect our tacit semantic intentions with regard to the content of our word 'water'.[12] There is a sense in which our semantic intentions determine which concept is expressed by our word 'water': they ensure that it expresses a concept that applies only to H_2O if water is actually H_2O, a concept that applies to anything water-like if water turns out to be highly disuniform, and so forth. But there is another sense in which our semantic intentions do not, by themselves, determine the content of 'water'. They leave it up to the environment to determine which of the default conditionals is the operative default conditional, that is, which has its antecedent satisfied. This is what makes the indicated default-conditional profile for 'water' a characteristically externalist profile. An internalist profile would not allow environmental variations to affect the content of 'water'; the consequent of each default conditional in its profile would be exactly the same, regardless of the antecedents.

These points generalize. For each kind term, there is a default conditional truth that specifies the nature of the concept that is expressed by the term if it turns out that the term fails to denote. In each case, the content of this default conditional is determined by our semantic intentions, and in each case its content can be discovered by means of a priori reflection. These default conditionals for the vacuous case may, in principle, differ fundamentally from one kind term to the next—but, evidently, there is no way that those default conditionals can turn out so as to be inconsistent with externalism, either in letter or spirit.

II. SEGAL'S DILEMMA

We are now in a position to see how an externalist might respond to Segal's argument from dry earth. Segal argues that the externalist can provide no plausible account of the application conditions of the concept expressed by the Dry English word 'water' (hereafter: the concept dwater). The argument takes the form of a dilemma: the concept dwater either has motley application conditions—where a concept has motley application conditions just in case it applies, actually and counterfactually, to substances on the basis of their exemplifying some special cluster of superficial properties—or else the concept is necessarily vacuous (*op. cit.*, pp. 54–56).[13] Segal argues that neither horn is available to the externalist. I will consider each in turn.

---

[12] The particular default conditionals mentioned in the text evidently cannot themselves serve as content-fixing intentions, since the word 'water' (whose semantic properties are to be fixed by these intentions) appears in the conditionals. But closely related conditionals can do the job, where 'water' is replaced by some expression along the lines of 'these samples' or 'clear, odorless, potable,... liquid'.

First Horn. On the first horn of Segal's dilemma, the externalist contends that the concept dwater has motley application conditions. This is the verdict that, according to Segal, accords with our intuitions about the concept dwater (*op. cit.*, p. 55). But he maintains that externalists cannot occupy this horn of the dilemma, for the following reason:

> The trouble is that it would be ad hoc for an externalist to allow this. It is the [second horn] that conservatively extends the normal externalist line on nonempty concepts to empty ones (*op. cit.*, p. 55).

Segal claims that a conservative extension of "the normal externalist line on nonempty concepts" must take the concept dwater to be necessarily vacuous. The trouble with Segal's objection is that there is no normal externalist line on nonempty concepts. While there is a normal externalist line on natural-kind concepts, externalism is silent with regard to non-natural-kind concepts. Externalists, for instance, are free to hold that such non-natural-kind concepts as food and shelter are motley concepts, which intrinsic duplicates cannot fail to share. Surely Segal does not expect externalists to extend the normal externalist line on natural-kind concepts to *all* non-natural-kind concepts. Empty concepts are themselves non-natural-kind concepts. Natural kinds, recall, are kinds, all of whose actual instances share some distinctive basic physical constitution; so empty concepts trivially fail to qualify as natural-kind concepts. Segal's insistence on a conservative extension of the normal externalist line on natural-kind concepts to all empty concepts seems no more compulsory than an extension of the normal externalist line on natural-kind concepts to all nonempty concepts.

(Notice also that, just as there is no single "externalist line" on nonempty concepts, there need be no unified treatment of empty concepts either. Many externalists have the intuition that it is metaphysically impossible that there be unicorns (given that there are none) and conclude, reasonably enough, that the concept unicorn is necessarily vacuous. But it is not a *consequence* of externalism that this concept is necessarily vacuous, nor that any other concept with an empty extension is necessarily vacuous. Furthermore, the intuition that it is impossible that there be unicorns gives one little reason to believe that it is impossible that there be (say) hoverboards, particularly if one additionally has the intuition that it is possible that hoverboards exist.[14] There is no more reason for externalists to treat all empty concepts as nonmotley than there is to treat all nonempty concepts as nonmotley. Nothing internal to externalism prevents externalists from honoring all of our intuitions about the application conditions of empty concepts, case by case.)

---

[13] These options do not exhaust the possible responses open to the externalist, but it is unlikely that he will opt for an alternative answer. For instance, the response that the concept dwater necessarily applies only to things of some one natural kind not only lacks prima facie plausibility but also evidently runs afoul of externalist thesis (E2). Externalists can, perhaps, respond that it is indeterminate whether the dry earth concept has motley or nonmotley application conditions. I shall not pursue this line of response here.

Those sympathetic with Segal may still feel uneasy about the disparate treatment of the concept water and the concept dwater, and they may contend that it is ad hoc for externalists to hold that the concept dwater has motley application conditions while maintaining that the concept water has nonmotley application conditions. We have seen that the "normal externalist line" does not require one to deny that the concept dwater is motley; but what reason can the externalist have for *affirming* that the concept is motley? I shall show that this horn can be independently motivated on the basis of the intuition that 'water' expresses a motley concept if (despite appearances) there is no water. Not everyone shares this intuition, and I address this intuitional conflict below in my discussion of the second horn of Segal's dilemma. For now, let us suppose that the intuitional conflict is settled in Segal's favor, that is, in favor of the claim that 'water' expresses a motley concept if there is no water. As we shall see, Segal's argument fails no matter how the intuitional conflict is resolved.

Recall that the inhabitants of dry earth are our intrinsic duplicates. Since water is a natural kind, and since there is no water on dry earth, externalist thesis (E2) entails that the inhabitants of dry earth do not have the concept water and, accordingly, do not have any beliefs or other mental states about water. But they do nevertheless have counterparts of these mental states. We believe that water is a liquid, and they believe that dwater is a liquid. When one of us is thinking that water is compositionally uniform, his duplicate on dry earth is having the thought that dwater is compositionally uniform. The same holds for our default-conditional beliefs and intuitions. We believe and intuit that if water is compositionally uniform then the concept expressed by our word 'water' necessarily applies only to substances that have that composition. Our dry-earth counterparts believe and intuit the counterpart of this default conditional: that if dwater is compositionally uniform then the concept expressed by the (Dry English) word 'water' necessarily applies only to substances that have that composition.

---

[14] Notice that Kripke's reasons for denying that there could have been unicorns (*Naming and Necessity* (Cambridge: Harvard, 1980), at pp. 156–58) seem not to extend to fictional artifactual kinds.

Finally, since they are our intrinsic duplicates, they will believe the counterpart of default conditional (iii) as well. They will believe that if (despite appearances) there is no dwater, then anything with the superficial qualities that they took dwater to have would be dwater. They would have the intuition that, if (despite appearances) there is no dwater, then the XYZ on a twin earth would be dwater. This all points toward their having a tacit semantic intention that their word 'water' behave in this way. And since there indeed is no dwater on dry earth, we would then have good reason to believe that their word 'water' does express a motley concept—for there is no reason to think that their semantic intentions are any less efficacious than our own.

Second Horn. On the second horn, the externalist contends that the concept dwater is necessarily vacuous.[15] Segal claims that this response is "plainly counterintuitive" (*op. cit.*, p. 55). (He adds that "[a]lthough there are no ghosts, there might have been." But, as we have seen, the concession that some empty concepts are unsatisfiable does not require the externalist to hold that all empty concepts (for example, hoverboard) are unsatisfiable. So Segal's observation that there could have been ghosts has no immediate bearing on the application conditions of the concept dwater.) There are two ways in which externalists might respond to the charge that this is counterintuitive.

First, externalists might concede that, intuitively, the concept dwater is satisfiable, but deny the evidential force of intuitions that tell against the necessary vacuity of this concept. This strategy threatens to leave externalists in a dialectically unstable position, since the whole of their evidence for externalism comes from twin-earth intuitions of just this sort. Those externalists who take intuition seriously are therefore best advised not to pursue this line of response. However, those externalists who hold that there is something other than intuitions that can justify our beliefs about twin earth will not be moved by this worry. Nor will those who accept a cost-benefit methodology on which one who relies evidentially on intuition is entitled to ignore those intuitions that tell against one's view so long as the theoretical payoffs of the view are sufficiently high.

---

[15] See, for instance, Tom Stoneham, "Boghossian on Empty Natural Kind Concepts," *Proceedings of the Aristotelian Society*, XCIX (1999): 119–22, and Sanford Goldberg, "An Anti-Individualist Semantics for Empty Natural Kind Terms," *Grazer Philosophische Studien*, LXX (2006): 55–76.

Second, externalists might respond that, intuitively, the concept dwater is necessarily vacuous. They might contend that, intuitively, if (despite appearances) there is no water on earth, then a sample of XYZ on a twin earth would not be water. And they can then exploit the same reasoning I used above to argue that the inhabitants of dry earth intend for their word 'water' to express a necessarily vacuous concept should it turn out that in fact there is no clear, odorless, potable, ... liquid anywhere on their planet. They might insist that Segal simply has aberrant intuitions. Such prima facie intuitional conflicts are old hat to externalists. Saul Kripke showed how, in related cases, intuitional conflicts may be explained away as confusions of a certain sort. But, until the externalist produces a satisfactory rephrasal strategy (or some other account of intuitional error) that successfully deflates the competing intuitions, he remains in a stalemate with Segal, whose intuition is that the dry-earth concept applies to all water-like substances.[16]

I wish to take no stand on the outcome of the stalemate, except to announce my optimism that it can be resolved. If the stalemate is resolved in favor of the intuition that the concept dwater is necessarily vacuous, then there is a straightforward answer to Segal's dilemma, for Segal's only objection to adopting the second horn of the dilemma is that it is counterintuitive. If, on the other hand, the stalemate is resolved in favor of the intuition that the concept has motley application conditions, then (as we saw above) this intuition can be used to establish that the inhabitants of dry earth intend for their word 'water' to express a motley concept given their circumstances, which provides the externalist with the leverage to answer Segal's argument against adopting the first horn of the dilemma. Very little rides on just whose intuitions turn out to be correct; either way, Segal's dilemma fails.

III. BOGHOSSIAN'S DILEMMA

We come now to Boghossian's dry-earth argument against externalism. The concept dwater must either be atomic or complex; but, argues Boghossian, neither option is open to the externalist. I will again examine both horns of the dilemma and argue that there is nothing, in principle, to prevent the externalist from occupying the horn of his choice.

---

[16] See George Bealer, "The Origins of Modal Error," *Dialectica*, LVIII (2004): 11–42, at pp. 23–31, and my "Law Necessitarianism and the Importance of Being Intuitive," *Philosophical Quarterly*, LV (2005): 649–57, at pp. 654–55 for discussion of the constraints on satisfactory rephrasal strategies.

First Horn. Externalists who have the intuition that 'water' expresses a complex concept if (despite appearances) it turns out that there is no water on earth may opt for the "compound option," according to which the concept water is an atomic concept while the concept dwater is complex. Here is Boghossian's argument that externalists cannot accept the compound option:

> The compound option requires the externalist to say that one and the same word, with one and the same functional role, may express an atomic concept under one set of external circumstances and a compound decompositional concept under another set of external circumstances.... But it is hard to see how the *compositionality* of a concept could be a function of its external circumstances in this way. Compositionality, as I understand it, can only be a function of the internal syntax of a concept; it can't supervene on external circumstances in the way that the compound proposal would require (*op. cit.*, pp. 172–73).

This "internal syntax" presumably is meant to be something that is duplicated in intrinsic duplicates (for example, some neurophysiological feature). So we may understand Boghossian as rejecting the compound option on the grounds that it violates the principle that the compositionality of one's concepts supervenes on one's intrinsic properties—that is, necessarily, if two individuals share all of their intrinsic properties, then the concept that the first employs at some time *t* is atomic (complex) iff the concept that the second employs at *t* is atomic (complex).

Boghossian's supervenience principle should not be confused with the more familiar internalist supervenience principle that the content of one's mental states supervenes on one's intrinsic properties. Like this more familiar supervenience principle, Boghossian's supervenience principle enjoys a high degree of initial plausibility. Also like this more familiar principle, Boghossian's supervenience principle appears to be open to counterexamples. I will give two.

The first counterexample is a variant of Tyler Burge's famous arthritis example.[17] The concept H_2O is a complex concept.[18] Suppose that Oscar acquires the word 'H_2O', understanding that it is a name for water, but without realizing its complexity (hearing it, perhaps, as

---

[17] I owe this example to Derek Ball. For Burge's original arthritis example, see his "Individualism and the Mental," *Midwest Studies in Philosophy*, IV (1979): 73–121, at pp. 77–79.

[18] Specifically, it is the complex concept of being dihydrogen monoxide.

"aichtuwo"). For the very same reasons that led us to conclude that Burge's arthritis-man has (but misunderstands) the concept arthritis, as opposed to having an idiosyncratic concept, we ought to conclude that Oscar nonetheless shares our (complex) concept H_2O. Twoscar—Oscar's intrinsic duplicate—is a member of a very different linguistic community in which 'H_2O' expresses an atomic concept: the experts, let us suppose, do not treat the 'H', '2', and 'O' as individually meaningful.[19] So, although they are intrinsic duplicates, 'H_2O' expresses a complex concept in Oscar's mouth and an atomic concept in the mouth of his duplicate. Hence, the complexity of one's concepts does not supervene on one's intrinsic properties alone.[20] One might object that this begs the question against Boghossian, since he would reject the social externalism that I take for granted in setting out the counterexample. But that is a mistake: this counterexample is no less dialectically appropriate than Burge's own counterexamples to individualism about mental content.

Here is the second counterexample. Suppose that my duplicate and I each introduce a word '*F*' with the following stipulation: Let '*F*' express the atomic concept *two* iff there are aliens and the complex concept *the even prime* iff there are no aliens. In the actual world (let us suppose) there are aliens. My duplicate inhabits a world otherwise identical to the actual world except that there are no aliens. It would be a mistake to think that we now express one and the same concept when we say '*F*'—to think, for instance, that we both express the concept *two iff there are aliens and the even prime iff there are not.* On the contrary, the nature of the stipulation guarantees that different concepts are expressed: '*F*' expresses the atomic concept *two* in my mouth and the complex concept *the even prime* in the mouth of my duplicate. Hence, the complexity of one's concepts does not supervene on one's intrinsic properties alone.[21]

---

[19] We may suppose that the experts in Twoscar's world have not yet discovered the periodic table and do not have a systematic nomenclature for chemical kinds. They introduce words like 'H_2O' to sound smart.

[20] Boghossian can (but presumably will not, given what he says about external circumstances) avoid the counterexample by abandoning the original supervenience principle and reverting to a weaker principle according to which individuals cannot differ with regard to the complexity of their concepts without some difference with regard to their intrinsic states *or* with regard to their respective linguistic communities. The compound option also runs afoul of this global supervenience principle, to which the Oscar case cannot serve as a counterexample. The second counterexample, however, is designed to refute both the local and the global version of the supervenience principle (for we can suppose that the entire community is present for the indicated baptism).

[21] This counterexample, like the first, crucially involves individuals who evidently do not fully understand the concepts that they employ. Both counterexamples can

The kind of content-fixing stipulation at work in this counterexample might raise some eyebrows, even among externalists.[22] But if Kripke is right that I can successfully fix the semantic properties of a name for some unfamiliar mineral simply by saying "Let '*G*' refer to stuff like this," I should likewise be able to fix its semantic properties with a more explicit baptism: "If this stuff is all SiO_2 then let '*G*' express a concept that applies only to SiO_2, and if this stuff is all Fe_2O_3 then let '*G*' express a concept that applies only to Fe_2O_3, and if ...." I see no reason for taking the baptism of *F* to be any less legitimate than this explicit baptism of *G*—if anything, the former is more secure, for more is known about the nature of *F* at the time of the baptism than about the nature of *G*.

The failure of this supervenience principle opens the door for the externalist to defend the compound option in much the same way that the first horn of Segal's dilemma was defended above. Just as an externalist is free to hold that it is built right into the default conditionals that external conditions are able to affect whether a motley or nonmotley concept is expressed, an externalist may likewise hold that the default conditionals encode different conceptual complexities on different empirical suppositions: if it turns out that water is compositionally uniform then 'water' expresses an atomic concept, whereas if it turns out that there is no water then 'water' expresses a complex concept. If this is right then, by exploiting the reasoning employed in section II, he can establish that the semantic intentions of the inhabitants of dry earth guarantee that their word

---

therefore be blocked by restricting the original supervenience principle to concepts that are fully understood by the individuals in question: for intrinsic duplicates *A* and *B*, if *A* fully understands *c* and B fully understands the counterpart concept *c*′, then *c* is atomic iff *c*′ is atomic. But the compound option does not run afoul of this new principle. For suppose that the compound option is correct, and suppose (as seems plausible) that one cannot fully understand a concept unless one knows, or is in a position to know, its complexity. Since we know that water has a uniform chemical composition, we are in a position to know that the concept water is atomic. We fully understand our concept water. But our duplicates on dry earth are not in a position to know that their concept dwater is complex, for they are under the mistaken impression that dwater is a natural kind. So they do not fully understand their concept dwater. Since they do not fully understand the concept, the compound option is consistent with the revised supervenience principle.

[22] See, for instance, Scott Soames, *Philosophical Analysis in the Twentieth Century*, Volume 2 (Princeton: University Press, 2003), at pp. 410–13. There are many moving parts in Soames's argument, and a proper discussion would take us too far afield. See Robin Jeshion, "Soames on Descriptive Reference-Fixing," *Philosophical Issues*, XVI (2006): 120–40, for a critical discussion of Soames's argument.

'water' expresses a complex concept should it turn out that there is no dwater.

I suspect that externalists will prefer the second horn for, among other things, many externalists (myself included) lack the intuition that 'water' expresses a complex concept if there is no water, in which case there is no independent reason to accept the compound option. But as far as I can see, the compound option has not been refuted.

Second Horn. On the second horn, the externalist maintains that the concept dwater is an atomic concept. Boghossian argues that the atomic option is unavailable to externalists because externalists lack the resources to say *which* atomic concept is expressed by the Dry English word 'water'. According to Boghossian, externalists cannot supply satisfactory answers the following two questions: (1) "What are the satisfaction conditions for 'water' on Dry Earth, to what sorts of liquid does it apply?" and (2) "What proposition...is expressed by sentences of the form 'Water is wet'... as uttered on Dry Earth?" (*op. cit.*, p. 173). But these questions can be answered.

Concerning the first question, externalism is silent with regard to the satisfaction conditions (that is, application conditions) of non-natural-kind concepts—so externalists may, consistently with externalism, say either that the concept dwater has motley satisfaction conditions or, alternatively, that it is necessarily vacuous.[23] Concerning the second question, the proposition expressed by the Dry English sentence 'Water is wet' is the proposition that dwater is wet. Finally, following Boghossian in taking propositions to be "composed of" the referents of the terms that occur in sentences that express them, the proposition that dwater is wet is "composed of" the property of being dwater, the property of being wet, and the instantiation relation.[24]

Why was the externalist supposed to be unable to answer these questions? Boghossian claims that externalists have "an overriding commitment to individuating a concept in terms of its referent" (*op. cit.*, p. 173), from which it is supposed to follow that "[s]ince there is no natural kind at the end of the relevant causal chain leading up to uses of 'water' on Dry Earth, there is no fact of the matter what the referent of 'water' is" (*op. cit.*, p. 174). But externalists need not—and, as far as I know, do not—accept that concepts are all individuated in terms of their referents. As I have stressed throughout, externalism is silent with regard to non-natural-kind concepts, such as jade, food, and dwater. It is thus open to the externalist to hold that non-natural-kind concepts, and in particular empty concepts, are individuated in some entirely different way from natural-kind concepts.

---

[23] The concept dwater, if it is an atomic concept, may nevertheless be intensionally equivalent to the complex concept of being a clear, odorless, potable, ... liquid. Intensional equivalence plainly does not suffice for sameness of complexity: the atomic concept water, after all, is intensionally equivalent to the complex concept of being the actual clear, odorless, potable, ... liquid.

[24] See McLaughlin and Tye (pp. 308–09) for further development of this line of response.

IV. APPLICATIONS

Privileged Access. One immediate result of my treatment of dry earth is that it blocks Boghossian's argument that externalism is incompatible with privileged access. The argument runs as follows: If externalism is correct then, if there is no water (that is, if earth is dry earth), then one cannot have the concept water, for it can be neither an atomic concept nor a nonatomic concept. So externalists must hold that it can be known a priori that, if one has the concept water, then water exists. Since one can know by introspection alone that one has the concept water, it follows that one can come to know that water exists by means of introspection and a priori reflection alone. But this cannot be known by means of introspection and a priori reflection alone. So externalists must deny that one can know by introspection alone whether one has the concept water.[25]

I have shown that Boghossian's argument fails: it is open to externalists to hold that one can possess the concept water even if there turns out not to be any water. Consequently, one cannot rule out the possibility that there is no water given only the introspective knowledge that one has the concept water. The argument for the incompatibility of externalism and privileged access can therefore be blocked. This is not a new solution to Boghossian's incompatibility argument—Brian McLaughlin and Michael Tye (*op. cit.*) and Sanford Goldberg (*op. cit.*) have argued that externalists are free to hold that the Dry English word 'water' expresses an atomic concept. My contribution has been to show that this account of the complexity of the concept is not mandatory (nor for that matter is any particular account of the application conditions of the concept) and to show how our semantic intentions and the default conditionals that encode them might underwrite various externalist accounts of the application conditions and complexity of this concept.

There is, however, a related thesis that externalists are committed to denying, if they are to accept the default conditionals specified in section I. For, in that case, one is not always in a position to know, solely on the basis of introspection and a priori reflection, whether a given concept has motley or nonmotley application conditions. Whether one is employing a motley or nonmotley concept will depend, in part, upon whether the kind in question is a natural kind, which is an empirical matter. Likewise, if one opts for the first horn of Boghossian's dilemma, then one will have to abandon the thesis that one is always in a position to know, solely on the basis of introspection and a priori reflection, whether a given concept is atomic or complex. Abandoning these particular privileged access theses is, however, entirely compatible with retaining the most basic privileged access thesis that, for all concepts *c* (and thoughts that *p*) if one has concept *c* (or thinks that *p*), then one can know that one has concept *c* (thinks that *p*) on the basis of introspection alone. And externalists have long ago abandoned the thought that all of a concept's properties (for example, its application conditions) can be known solely on a priori and introspective grounds.

---

[25] Some externalists may be willing to bite the bullet here and deny that one can know by introspection alone whether one has the concept water; see, for example, Andrew Cullison, "Privileged Access, Externalism, and Ways of Believing," *Philosophical Studies*, forthcoming.

Language of Thought. Proponents of the Language of Thought hypothesis hold that concepts are symbols in a mental language. They standardly also accept the auxiliary thesis that facts about the (syntactic) complexity of the symbols that one tokens supervene on physical features of those tokens—features that do not vary across intrinsic duplicates. In section III, I presented two arguments that intrinsic duplicates can differ with respect to the complexity of the concepts that they employ. If either of those arguments is successful, then proponents of the Language of Thought hypothesis must abandon this auxiliary, individualistic assumption.

Empty Names. Peter Ludlow has recently defended a mixed theory of names on which ordinary proper names refer directly while fictional names have descriptive content.[26] This view promises to secure all of the advantages of both the Millian treatment of ordinary proper names and the Russellian treatment of empty names. Perhaps the familiar Kripkean arguments against descriptivism can be adapted to show that fictional names lack descriptive content and that the mixed view is therefore untenable.[27] But there are various other arguments against the mixed view that can be answered in light of the points raised above about the semantics of natural-kind terms.

---

[26] See his "Externalism, Logical Form, and Linguistic Intentions," in Alex Barber, ed., *The Epistemology of Language* (New York: Oxford, 2003), pp. 399–414, on pp. 404–09.

[27] It does not seem possible to Gödel-Schmidt Santa Claus. The modal arguments, however, appear to retain much of their force. For further criticism of the mixed view, see Marga Reimer, "A 'Meinongian' Solution to a Millian Problem," *American Philosophical Quarterly*, XXXVIII (2001): 233–48, and David Braun, "Empty Names," *Nous*, XXVIII (1993): 449–69.

First, one might object to the mixed view on the grounds that it permits intrinsic duplicates to differ with respect to the complexity of their concepts. For instance, I have a duplicate in a world in which (unbeknownst to him) the counterparts of our Sherlock Holmes stories document the activities of a real individual. Accordingly, on the mixed view, my name 'Sherlock Holmes' has a descriptive sense while my duplicate's sound-alike name refers directly. The counterexamples to Boghossian's supervenience principle in section III can be adapted to provide independent support for the possibility of this kind of divergence. For instance, 'Princess Diana' is a partially descriptive name; even direct reference theorists will agree that it is semantically associated with a descriptive property.[28] I have a duplicate in a world in which Diana's counterpart is not a princess, but in which she adopts the name 'Princess Diana' as a stage name (like 'Queen Latifah'). My duplicate, like myself, wields his term 'Princess Diana' deferentially; and since, by Millian lights, 'Princess Diana' lacks descriptive content in the mouths of those to whom he defers, it lacks descriptive content in his own mouth as well. So, despite being intrinsic duplicates, it is not the case that his terms have descriptive content just in case mine do.

Second, one might object to the mixed view on the grounds that it provides a disuniform account of the semantics of names. But, as we saw in the case of natural-kind terms, even when our semantic intentions underwrite disuniformities in the application conditions or complexities of the associated concepts, uniformity reemerges at the level of the default conditionals. Suppose, for instance, that a priori reflection reveals that if in fact there is no water, then 'water' expresses a complex concept with motley application conditions. In that case, despite their semantic differences, the English and Dry English words 'water' are both meant to express an atomic nonmotley concept if it turns out that the relevant substance is a natural kind and are both meant to express a complex motley concept if it turns out that there is no such substance. Likewise, although 'George Washington' and 'Sherlock Holmes' are assigned fundamentally different semantic values on the mixed view, uniformity is secured at a deeper level insofar as both are intended to directly refer if an appropriate individual stands at the end of the relevant naming chain, and to have a descriptive sense otherwise.

DANIEL Z. KORMAN

University of Texas at Austin

---

[28] See Soames, *Beyond Rigidity* (New York: Oxford, 2002), at pp. 110–30.

DISJUNCTIVE CAUSES*

There is an initial presumption against disjunctive causes. First of all, for some people causation is a relation between events. But, arguably, there are no disjunctive events, since events are particulars and thus they have spatiotemporal locations, while it is unclear what the spatiotemporal location of a disjunctive event could be.[1] More importantly, even if one believes that entities like facts can enter in causal relations, and even if there are disjunctive facts, it is still hard to see how disjunctive facts could be causes. Imagine, for instance, the following scenario. I have a gun filled with red paint and another gun filled with blue paint, and I fire both guns at my neighbor's white wall. A moment later, there is graffiti on the wall and my neighbor notifies the police. He would have done so regardless of the graffiti's color, since all he cares about is the existence of graffiti on his wall. Is it plausible to claim that a disjunctive fact is a cause of his notifying the police? In particular, is it plausible to claim that he notified the police because I fired the red-paint gun *or* the blue-paint gun (the thought being that my firing paint of either color *would have sufficed*)? It seems not. The police were notified because of the actual graffiti on the wall, and the actual graffiti on the wall is made of a certain pattern of colored patches. Imagine, that, as it turns out, there are patches of both colors on the wall. Then it seems that both my firing the red-paint gun and my firing the blue-paint gun were causes of my neighbor's notifying the police. In other words, my firing the red-paint gun and my firing the blue-paint gun jointly caused the outcome: each of them was a contributory cause of the outcome's occurrence. On the other hand, imagine that there are only patches of one color on the wall. Then it seems that my firing only one of the guns was a cause. Either way, the disjunction fails to be a cause: either my firing the red-paint gun was a cause, or my firing the blue-paint gun was a cause, or they were both causes, but their disjunction was

---

* For valuable discussion and comments, thanks to Ellery Eells, Martha Gibson, Daniel Hausman, Russ Shafer-Landau, Lawrence Shapiro, Alan Sidelle, Stephen Yablo, the audience at the 2005 Pacific American Philosophical Association (in particular, my commentator, Brad Armendt), the members of the metaphysics and epistemology reading group at the University of Wisconsin-Madison, two editors of this JOURNAL, and especially Juan Comesaña.

[1] For an argument of this type against disjunctive events, see David Lewis, "Events," in *Philosophical Papers, Volume II* (New York: Oxford, 1986), pp. 241–69.

0022-362X/06/0310/521–38

not. This seems to generalize: for every outcome that could have been caused by more than one route, the relevant disjunctive fact (the fact that *some* route was active) is not a cause of the outcome; instead, the outcome's causes are determined by the properties of the actual process issuing in the outcome's occurrence. Thus, it seems that causes cannot be disjunctive.

Now, in a few exceptional cases, some people are prepared to drop this presumption and to accept the existence of disjunctive causes. Two main types of case come to mind. First, overdetermination cases: Suppose that, improbably enough, two rocks hit a window at exactly the same time and the window shatters as a result. Suppose that each rock was independently sufficient for the shattering. What caused the shattering? If we were to say that both rocks did, we would be committed to a superabundance of causes (more causes than are needed to account for the fact that the window shattered). By contrast, if we were to say that the shattering was caused by the fact that *at least one* of the rocks struck the window—a disjunctive fact—the redundancy would be avoided. As a result, some people have thought that we should say that the disjunctive fact is a cause.[2] A second type of case in which someone might be prepared to accept the existence of disjunctive causes is causation by omission (or by absences in general). Omissions seem to be overly disjunctive in nature: they obtain just in case the world is one out of many different possible ways. For instance, my failure to go to the movies obtains just in case I stay at home or I go to a party or I go for a walk, and so on. Thus, if omissions are causes, and if omissions are disjunctive, it follows that there are some disjunctive causes.[3]

Now, what we should say about cases involving overdeterminers and omissions is controversial. For one thing, some people find that there is nothing problematic about causal redundancy. If so, there is no motivation to posit the existence of disjunctive causes in cases of overdetermination. As for omissions, some people simply reject the possibility of their entering in causal relations. Still others accept the possibility of causation by omission but reject the claim that omissions are disjunctive (for example, by identifying omissions with commissions). In any case the commitment to disjunctive causes is avoided.[4]

---

[2] See, for example, Jonathan Bennett, *Events and Their Names* (Indianapolis: Hackett, 1988); John Mackie, "Causes and Conditions," in Ernest Sosa and Michael Tooley, eds., *Causation* (New York: Oxford, 1993), pp. 33–55; D.H. Mellor, *The Facts of Causation* (New York: Routledge, 1995).

[3] David Lewis discusses this view on omissions in Postscript D to "Causation," in *Philosophical Papers, Volume II* (New York: Oxford, 1986), pp. 189–93. Other cases where one might be led to think that disjunctions are causes are "preemptive prevention" cases (these cases also involve absences, although as intermediaries in the chain; see note 16 below).

At any rate, setting aside exceptional and controversial cases like those involving overdeterminers and omissions, the presumption that causes are not disjunctive is quite general. In other words, the common thought is that, at least in ordinary circumstances, disjunctive facts cannot be causes. And a reason to believe this is that, as the graffiti case illustrates, it is hard to see how causes could be disjunctive in ordinary circumstances: it seems that, regardless of what the world is like, disjunctive facts cannot causally contribute to the occurrence of outcomes.[5]

In what follows I argue that this impression is misguided. I suggest that there is an intelligible way the world could be like for disjunctive facts to be causes in ordinary circumstances, and that the view that disjunctions are causes in those circumstances is supported by some intuitions about causation. I lay out the argument for disjunctive causes in the next section. Then, in section II, I explain how the argument can be generalized and, in section III, I examine some of its potential implications.

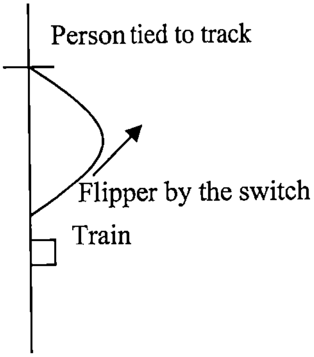
I

Let us start by imagining the following scenario. A runaway train is running on a track, where a person is trapped, up ahead. There is a switch and a side track. Someone flips the switch and as a result the train turns onto the side track—call this person *Flipper.* Unfortunately, the tracks reconverge after a while, where the person is trapped, and thus he still dies (in pretty much the same way, at around the same time). The situation is depicted by the following picture:

---

[4] For an argument that overdetermination is widespread, see Jonathan Schaffer, "Overdetermining Causes," *Philosophical Studies,* CXIV (2003): 23–45. For an argument that omissions cannot enter in causal relations, see Helen Beebee, "Causing and Nothingness," in John Collins, Ned Hall, and Laurie Paul, eds. *Causation and Counterfactuals* (Cambridge: MIT, 2004), pp. 291–308.

[5] Alan Penczek appears to be an exception—see his "Disjunctive Properties and Causal Efficacy," *Philosophical Studies,* LXXVI (1997): 203–19. Penczek seems to believe that any ordinary case where an outcome could be caused in one of two ways might be a case where the outcome has a disjunctive cause. But I find his argument unconvincing. Penczek would argue, for instance, that if I shoot a bullet that kills a person but there was a backup assassin waiting in reserve, then my shooting is not a cause because the relevant counterfactual dependencies fail to obtain (in particular, the death would still have occurred if I had not shot). Thus, since we do not want to say that the outcome is uncaused, we should say that a disjunctive condition (presumably, the fact that at least one of us shot) is the cause. However, if anything, cases of this type are a reason to believe that counterfactual accounts of causation are wrong, not a reason to believe in disjunctive causes.



What were the causes of the person's death in this case? Clearly, we would say that they include: the train starting to run on the tracks, the driver's becoming unconscious, the person's foot getting stuck in the tracks, the train hitting the person, and so on. On reflection, we would probably also identify other things as causes, in particular, events or conditions without which the death would not have occurred, but that are normally less salient to us, such as the tracks being in working order or the train's not falling apart as it travels down the track. All these, philosophers tend to think nowadays, are causes, although the contribution of the latter (which we could refer to as "background conditions") is sometimes unnoticed because it is too obvious to be worth mentioning.

Now, how about Flipper's action? Is his redirecting the train a cause of the person's death? On the face of it, it is not. It is clearly not one of the salient causes. Also, it does not seem to be one of the background conditions: it does not contribute something towards the outcome that is too unexciting or obvious to be worth mentioning. What the switching does is help determine the actual *route* to the outcome, and, intuitively, determining the route to an outcome is not tantamount to making a causal contribution to the occurrence of the outcome itself (however boring or obvious). To illustrate with another example: we feel that it would be wrong to say that an air stream that temporarily deflects a missile from its path (where the missile then resumes its original path) causes the target's destruction. This is so even if we think that the destruction of the target occurred due to many contributing causes (such as the missile's being launched, someone's giving the order to launch it, the presence of the target, the missile's never losing the necessary momentum to reach the target, and so on). For we do not think that the deflecting air stream was one of those contributory causes.

The judgment that Flipper's action is not among the causes of the death is quite generally recognized in the philosophical literature on causation and there have been several attempts to capture it.[6] Recently, however, Ned Hall has argued that such judgment is misguided.[7] Although this is not the place to offer a full defense of the intuition from possible objections (my main goal here is to show that an argument for disjunctive causes can be built on that basis, *if* the intuition is to be trusted), I do think that we should aim to capture that intuition, and thus I will quickly review my main reasons for thinking this.

First, as I have already pointed out, the contribution of a switching event (like Flipper's redirecting the train) seems to be importantly different from even that of the least salient causes.[8] Second, the intuition that switches are not causes seems to be importantly tied to the idea that causes are "difference-makers," which is a fundamental idea about causation—in particular, the reason Flipper appears not to be a cause is that it "fails to make a difference" in the relevant sense.[9] And finally, the intuition that switches are not causes can do important theoretical work in grounding certain *moral* claims that we want to make about switching scenarios. For instance, we want to say that Flipper is not morally responsible for the person's death. Why is he not morally responsible? Intuitively, because he is not a cause of the death. Thus preserving the causal intuition allows us to ground attributions of moral responsibility on attributions of causal responsibility in the natural way.[10]

---

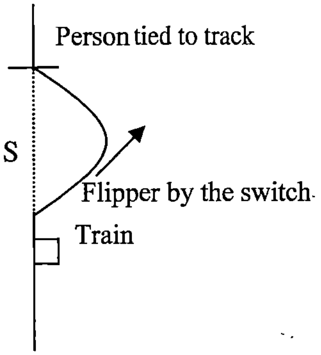
[6] See William Rowe, "Causing and Being Responsible for What Is Inevitable," *American Philosophical Quarterly*, XXVI, 2 (1989): 153–59; Laurie Paul, "Aspect Causation," this JOURNAL, XCVII, 4 (April 2000): 235–56; Stephen Yablo, "De Facto Dependence," this JOURNAL, XCIX, 3 (March 2002): 130–48; Carolina Sartorio, "Causes as Difference-Makers," *Philosophical Studies*, CXXIII, 1–2 (2005): 71–96.

[7] Hall, "Causation and the Price of Transitivity," this JOURNAL, XCVII, 4 (April 2000): 198–221 (Hall calls cases of this type "interactive switches," given the existence of a physical connection between the switching event and the outcome). Hall seems to now have changed his view on these matters (personal communication).

[8] In "Causation and the Price of Transitivity," Hall argued that switching events are boring causes on a par with background conditions. His argument went as follows. The switching's contribution is like the contribution of the existence of the side track a day earlier (a boring cause). For the relation between the switching and the setting of the switch as the train passes over it is just like the relation between the existence of the side track the day before and its existence today, when the train travels down it; in addition, the relation between the setting of the switch as the train passes over it and the outcome is just like the relation between the existence of the side track today and the outcome. Hall seems to be thinking in terms of physical connectedness: just like we can trace a continuous physical process linking the side track's presence the day before to the outcome today (via the presence of the track when the train passes over it), we can trace a similar process linking the switching event to the outcome (via the setting of the switch when the train passes over it). However, unless we assume that physical connectedness is sufficient for causation (which is debatable), this fails to show that the two relationships are the same in all relevant respects. In particular, as I have pointed out, a main feature of the switching event is that what it does can be seen as the deflection of an ongoing process (in virtue of which the switching helps to determine the actual route to the outcome, but this is all it does). In contrast, the presence of the side track the day before plays no similar deflecting role.

[9] In what sense, exactly, does the switching fail to make a difference? One might think that it fails to make a difference in that the outcome would still have occurred in its absence; however, the existence of "preemption" cases shows that it is possible for an

Now, an important feature of the train case, as I am conceiving it, is that the person would still have died if Flipper had failed to flip the switch (and maybe also that the redirection does not significantly alter the properties of the death). If the person would not have died had Flipper not flipped the switch, then, presumably, flipping the switch *would* be a cause of the death. For instance, imagine that the stretch of the main track where the two tracks come apart (call it *S*) is *disconnected*, as depicted by the following picture:



---

event to be a cause of an outcome even if the outcome would still have occurred in the absence of the event. In my "Causes and Difference-Makers," I suggest that causes are difference-makers in this other sense: if an event is a cause, then its absence would not have been a cause of the same outcome. The switching fails to make a difference in this sense, for, if we were to count it as a cause, then, had it not occurred, we would have had to count its absence as a cause too.

[10] A clarification is in order: as we will see later (see section III below, and, in particular, note 17), I think that the fact that Flipper failed to cause the death can only be *part* of the explanation of his lack of moral responsibility; still, I think that it is an important part of the explanation. By the way, note that, although Flipper's lack of moral responsibility seems importantly connected with his alleged lack of causal responsibility, our intuition that he is not a cause is not "tainted" by the moral judgment, but it is a purely metaphysical intuition. For we still think that the redirection fails to cause the death if what redirects the train is, say, a gust of wind instead of a moral agent.

In this case, if Flipper had failed to flip the switch, the train would have derailed while passing through *S* and the person's life would have been spared. Thus flipping the switch *would* be a cause of the death. However, if the train would not have derailed if it had stayed on the main track because *S* is connected at the relevant time, then flipping the switch would not be a cause of the death. Let this be claim 1:

> C1: If S is connected and thus the train would have reached the person via the main track, then Flipper's action is not a cause of the person's death.[11]

Now set Flipper aside for a moment. Imagine that *S* has been temporarily disconnected for repairs, and that someone reconnects it when the train passes through the switch—call him *Reconnecter.* Is Reconnecter's action a cause of the death of the person trapped on the tracks? Clearly, it depends: it depends on whether or not the switch was pulled. If the switch was not pulled and the train stayed on the main track, then the reconnection of *S* seems to be a cause of the death because, had it not been for the reconnection, the train would have derailed and the death would not have occurred. On the other hand, if the switch was pulled, then the train never had a chance to run on *S.* In that case, clearly, the reconnection of *S* is *not* a cause of the death (I am imagining that the pulling of the switch and the reconnection of *S* were causally independent events: neither occurred because the other occurred). Let this be claim 2:

> C2: If the switch is (independently) pulled, and thus the train never runs on *S,* then Reconnecter's action is not a cause of the person's death.

Again, this intuition is very strong: if the train never ran on *S* (and what the train actually did was not affected by *S*'s reconnection in any way) then the reconnection had no effect on the person's fate.

On the basis of these claims, the argument for disjunctive causes is the following. Imagine a situation involving the train, the person tied to the track, and *both* Flipper and Reconnecter. In particular, imagine that, when the train is about to pass through the switch, Flipper flips the switch and Reconnecter independently reconnects *S.* The first premise of the argument is about Flipper. It says:

> (1) Flipper's action is not a cause of the death.

---

[11] Notice that I am not claiming that it is *generally* true that, when a person's death would still have occurred if one had acted differently, then one's action does not cause the death. The fact that the death would still have occurred if Flipper had not acted clearly is part of the explanation of why he fails to be a cause, but it is not the whole explanation (given the existence of preemption cases; more on this later).

This follows from C1: given that Reconnecter reconnected *S*, the train would also have reached the person if Flipper had not flipped the switch and the train had stayed on the main track. Thus, it follows from C1 that Flipper's action is not a cause of the death.[12]

The second premise is about Reconnecter. It says:

(2) Reconnecter's action is not a cause of the death.

This follows from C2: given that Flipper flipped the switch, the train never ran on *S*. Thus, C2 entails that Reconnecter's action is not a cause of the death.

Now, the third premise says:

(3) Flipper and Reconnecter *together* caused the person's death.

The justification for (3) is the following. If neither Flipper nor Reconnecter had acted the way they did, then the train would have stayed on the main track, it would have derailed while passing through *S*, and the person would have lived. Thus, the person died because of what the two of them did.

Now, what does it mean to say that Flipper and Reconnecter *together* caused the death? It could mean one of two things. First, it could mean that the death happened partly because of what Flipper did *and* partly because of what Reconnecter did. Second, it could mean that the death happened because *at least one of them* did what they did. The first suggestion is the "conjunctivist" suggestion, according to which the sum of the two actions, or a conjunctive fact, caused the death. The second suggestion is the "disjunctivist" suggestion, according to which a disjunctive fact caused the death. This is the fourth premise of the argument:

(4) If Flipper and Reconnecter together caused the death but their individual actions did not, then either the sum of the two actions (a conjunctive fact) caused the death or a disjunctive fact (the fact that either Flipper acted or Reconnecter acted) caused the death.

Finally, the fifth premise says:

(5) If either the sum or the disjunctive fact caused the death but the individual actions did not, then the disjunctive fact caused the death.

Note that, from (5) and the other premises, it follows that:

(6) There are disjunctive causes.

---

[12] One might think that the exact time of the reconnection matters to whether Flipper is a cause. I consider this possibility in section II.

Why believe in (5)? I will mention two reasons. First, there is counterfactual dependence between the disjunctive fact and the death, but not between the conjunctive fact and the death. There is counterfactual dependence between two actual events or facts *C* and *E* just in case, if *C* had not occurred/obtained, then *E* would not have occurred/obtained (or, according to the standard semantics for counterfactuals, if in the closest possible world(s) where *C* does not occur/obtain, *E* does not occur/obtain either). It is clear that the death counterfactually depends on the disjunctive fact: any nearby world where the disjunctive fact does not obtain is a world in which the person tied to the track lives. For it is a world where *both* Flipper and Reconnecter fail to act in the way they do in the actual world, and, as a result, in any such world the train continues to run on the main track and derails while passing through *S*. By contrast, arguably, the death does not counterfactually depend on the sum of Flipper's and Reconnecter's actions. There are nearby worlds in which the sum of the two actions does not obtain and in which the person tied to the track still dies. For instance, a world in which Flipper flips the switch but Reconnecter does not reconnect *S* is a world in which the sum of the two actions does not obtain (given that one of those actions fails to obtain) but the person still dies (in that world, the train runs on the side track for a while, then on the main track again, and finally kills the person). Furthermore, that world seems to be closer to actuality than a world in which *neither* of the two actions obtains, which differs from the actual world in what both Flipper and Reconnecter do. As a result, the death does not counterfactually depend on the sum of the two actions. So, if, as it is generally believed, the existence of counterfactual dependence is a strong indication of the existence of a causal relation, the disjunctive fact is a better candidate to be a cause than the sum of the two actions.[13]

---

[13] Note that the argument does not rest on the claim that counterfactual dependence *is* causation (which, as has often been pointed out, is most likely false) but only on the weaker claim that there is a strong *correlation* between counterfactual dependence and causation.

For an alternative view of the relevant notion of similarity among possible worlds, see David Lewis, Postscript E to "Causation," in *Philosophical Papers, Volume II* (New York: Oxford, 1986), pp. 211–12. There Lewis claims that the closest world to one where a sum *A&B* does not obtain is one where neither *A* nor *B* obtain, or, in other words, one where the event has been completely "excised" from the world, which includes removing all of its parts. But it is unclear that this is how we ordinarily evaluate counterfactuals. Recall that, in the case of Flipper and Reconnecter, the two actions are independent. Hence, imagining that one is absent should not lead us to imagine that the other one is absent too. At any rate, the point would remain that, whereas the outcome *clearly* depends on the disjunctive fact, this is not true of the conjunctive fact.

A second reason to prefer the disjunctivist view to the conjunctivist view is that it is in general implausible to attribute causal powers to a sum if the sum did not get those causal powers from the parts. The relation between a sum and its parts is mereological composition: a sum is composed of its parts just like ordinary objects (say, rocks) are composed of atoms. Now, we would not say that a rock causes something, say, the shattering of a window, unless we believed that its atoms (at least some of them) causally contributed to the shattering. Similarly, it seems wrong to suggest that a sum can cause something without any of the parts causally contributing to the same outcome. But this is what the conjunctivist view would have us say: according to the conjunctivist view, although Flipper's and Reconnecter's individual actions are not causes, their sum is. This problem does not arise—or, at least, it does not arise with the same force—in the case of disjunctive facts. For the relation between a disjunctive fact and its disjuncts is not composition: a disjunctive fact is not made "out of" its disjuncts in the same way that a rock is made out of the rock-atoms.

To illustrate this point further, it is helpful to consider the case of omissions again. Intuitively, some omissions are causes. For instance, intuitively, my failure to save a child who is drowning in a pond a few feet away from me is one of the causes of the child's death. Now, suppose for a moment that omissions are disjunctive facts; thus, my failure to save the child is my either staying on the shore eating ice cream or my walking home or... (so on and so forth). It is not implausible to suggest that the disjunctive fact (my omission) had causal powers that none of the disjuncts had. Intuitively, the child died because of what I failed to do, not because of what I did in its place. Suppose, for instance, that I was eating ice cream on the shore the whole time. It is plausible to suggest that it was not my eating ice cream *per se,* but only what it *entailed* (that is, the fact that I failed to jump into the water to rescue the child) that caused the child's death. Hence, there is no general motivation for believing that, when (if) a disjunctive fact is a cause, at least one of its disjuncts must also be a cause.

This concludes my argument for disjunctive causes. As we have seen, the argument is based on a scenario where two individual events (Flipper's action and Reconnecter's action) fail to be causes of a given outcome (as strongly suggested by intuitive judgments C1 and C2), but where the outcome happens as a result of the actions taking place together. In that case, the disjunctive fact is the best candidate for being a cause. In the next section I discuss how this argument can be generalized.

II

The basic fact that the case of Flipper and Reconnecter helps to bring out is the following. Many times there is more than one possible route to an outcome. Now, *sometimes*, when different routes to an outcome are simultaneously viable, a fact about one of the routes makes a fact about the other routes causally irrelevant. In particular, if one route is actualized at the same time that an obstacle to another route is removed, then the actualization of the former can render the clearing up of the latter irrelevant, and the clearing up of the latter can render the redirection to the former irrelevant. This is what happens in the case of Flipper and Reconnecter: the redirection to the side track makes the reconnection of the main track causally inefficacious, and, *vice versa,* the reconnection of the main track makes the redirection to the side track causally inefficacious. If either event had happened without the other, then that event would have been causally efficacious (if the reconnection had happened without the redirection, then the reconnection would have been a cause, and if the redirection had happened without the reconnection, then the redirection would have been a cause). But, when both events happen, they deprive each other of causal efficacy. Now, the argument continues, the outcome's occurring still depends on the existence of *some* viable route. Thus, the fact that some route was viable—a disjunctive fact—is causally relevant to the outcome. Hence, there are disjunctive causes.

To clarify: I do not mean to suggest that the argument for disjunctive causes generalizes to *every* case where there is more than one potential route to an outcome. This is clearly false. So-called "preemption" cases illustrate this: in a case of preemption, there are two possible routes to an outcome but one of those routes *preempts* the other, this is to say, it renders the other route causally inefficacious by exerting its own causal efficacy. For example, if two bullets are fired at a target and one of the bullets hits the target first, the first bullet preempts the second bullet. In this case, the outcome fails to have a disjunctive cause; rather, the first bullet is a cause and the second bullet is not.

Under what conditions, then, does the existence of more than one potential route to an outcome give rise to a disjunctive cause? The answer is: only in cases where there is an event that accounts for the switch in routes without, intuitively, being a cause of the outcome. The train scenario is a case of this sort: the two tracks are alternative routes to the outcome (the person's death) and, intuitively, all that the pulling of the switch does is to redirect an ongoing threat from one path to the other, without thereby being a cause of the outcome. In other words, the pulling of the switch determines the route to the death without being a cause of the death. In contrast with the train case, in a case of preemption, where, for example, I shoot the first bullet at the target and someone else shoots the second bullet that then soars through empty space, intuitively I do cause the outcome (I do not merely redirect a preexisting threat).

The contrast, then, is between "switching" cases and "preemption" cases. However this difference is to be accounted for (I will not attempt to do this here),[14] the suggestion is that the argument for disjunctive causes only generalizes to cases with the switching structure. The general form of the argument is as follows. Suppose that there are two (or more, but let us stick with two) potential routes to an outcome (say, *A* and *B*). Then, *in cases with the switching structure,* changing from *A* to *B* is not causally contributing to the outcome. Also, *A*'s being viable does not causally contribute to the outcome if *A* is not used. However, the outcome would not have occurred if the switch from *A* to *B* had not taken place and if *A* had not been viable. Hence, since the disjunctive fact is a better candidate cause than the sum of the two events, it (the fact that either there was a switch from *A* to *B* or *A* was viable) is a cause of the outcome's occurring. Therefore, the outcome has a disjunctive cause.

Now we can see where the initial presumption against disjunctive causes went wrong. As I have pointed out, that presumption is based on the impression that there is no way the world could be like for causes to be disjunctive (at least in ordinary circumstances). I illustrated this impression with the graffiti example: even if the police would have been notified had the graffiti been made out of blue or red paint (or both), there is no possible state of affairs in virtue of which the cause of the police being notified could be a disjunctive fact (for example, the fact that I fired a red-paint gun or a blue-paint gun at the wall). This is so because, if we look at the actual process issuing in the creation of the graffiti and thus in the police being notified (that is to say, if we follow the actual traces of blue and red paint), we will be able to determine that the cause was the firing of one gun (if there are only traces of one color) or the firing of both (if there are traces of both colors), but the fact that *either* gun was fired would never fill the bill.

However, this reasoning fails to generalize to cases with the switching structure. An important difference between the graffiti case and switching cases is the following. Imagine that, as it turns out, the graffiti is only made out of blue paint. In that case, the mere fact that the graffiti *could* also have been made out of red paint (in which case the police would still have been notified) is not sufficient to conclude that my firing the blue-paint gun was not a cause of the police being notified. In contrast, in the train case (and in switching scenarios in general), the mere fact that there was an alternative route is sufficient to rob the event of the redirection of its causal powers. Intuitively, the redirection is not a cause given that there was an alternative route, even if that route is never actualized. That is the distinguishing mark of a switching case: in a switching case, the status of a merely potential route can affect the causal status of the redirection to the actual route. As a result, there *can* be disjunctive causes. We just could not see it because we were focusing on cases with the wrong causal structure.

---

[14] This is the type of causal difference that the authors mentioned in note 6 aimed to capture.

Now, someone might worry that the Flipper and Reconnecter case has taught us nothing new. In particular, someone might argue that it is just another case of overdetermination, and that the argument from overdetermination cases to disjunctive causes is—as I have acknowledged in my introduction—one that has already been recognized. Why would one be tempted to think that the Flipper and Reconnecter case is a case of overdetermination? Because the switching and the reconnecting behave at least *like* overdeterminers in that, if each had taken place without the other, the outcome would still have occurred. If the reconnection had happened without the switching, or if the switching had happened without the reconnection, the death would still have occurred. Thus, one might think that the death was overdetermined by the two events.

However, the switching and the reconnecting differ from typical overdeterminers in at least two main ways. First, in a paradigm case of overdetermination (such as two rocks simultaneously hitting a window and making it shatter, or two bullets simultaneously entering a person's heart and causing that person's death), there are two easily distinguishable processes leading up to the outcome, and both of those processes are equally "active." In the train case, by contrast, we would not ordinarily say that there are two active processes, but only one: the actual trajectory of the train (if there had been two trains instead of one, each train running on one of the tracks, and if the two trains had reached the person at the same time, then that would have been a standard case of overdetermination). So, at the very least, the switching and the reconnecting are not *typical* overdeterminers. Whether in the end we should count them as overdeterminers is a separate matter, which I prefer to set aside here. The bottom line is that, even if they are overdeterminers, they are a special variety of overdeterminers, and as such they can supply a new set of reasons to believe in disjunctive causes.

Another important way in which the switching and the reconnecting differ from typical overdeterminers is that they need not occur simultaneously for them to give rise to the argument for disjunctive causes. Typical overdetermination cases involve some sort of unlikely coincidence at some point along the chain leading to the outcome (for instance, two rocks striking a window at exactly the same time, or two bullets entering a person's heart at exactly the same time). But this need not be the case with the switching and the reconnecting for the argument for disjunctive causes to work. To be sure, there are timing restrictions. In particular, one might argue that S had to be connected by the time the switch is actually pulled. Or one might argue that S had to be connected by the time the train *would* have reached it if the switch had not been pulled. Now, regardless of how exactly we should understand this restriction, it certainly does not require anything like a striking coincidence. In particular, it would be too strong to require that the reconnection and the switching happen at the same time. Suppose, for instance, that the reconnection happens a few seconds *before* the switching. In that case the argument for disjunctive causes would still have the same force: the reconnection would make the switching causally inefficacious and, in turn, the switching would make the reconnection causally inefficacious.[15]

---

[15] Recall that we are assuming that the switching and the reconnecting are causally independent events. How exactly to parse the timing restriction is an interesting problem. What if the reconnection happens *after* the switching? In his "De Facto Dependence," Yablo develops a theory according to which, very roughly, meeting a "need" is sufficient for being a cause. Yablo's theory seems to entail that the switching *is* a cause if the reconnection happens after the switching (although not otherwise) because then the switching meets a need for the death. But I am not sure that this is the right result. Suppose that there is a flawless mechanism already on its way to reconnect S, and it will do so before the train reaches S. It is somewhat implausible to suggest that the switching is only a cause if the mechanism reconnects S before the switching takes place. Clearly, whether the mechanism would operate in time for the train not to derail matters, but whether it would do so before or after the switch is flipped seems otherwise irrelevant.

Given that the case of Flipper and Reconnecter does not require an improbable coincidence, it would still be an "ordinary" case even if it were a case of overdetermination (in the sense I have used the word "ordinary" in my introduction). There might be other "atypical" overdetermination cases that do not involve improbable coincidences. For instance, election cases where a candidate wins by more than one vote: arguably, in those cases the outcome of the election is overdetermined but it does not involve any unlikely coincidence. Also, if a 250-pound man steps on a scale that only registers up to 200 pounds, one might argue that what the scale registers is overdetermined, but it does not involve any coincidence. (In his "Overdetermining Causes," Schaffer argues that overdetermination without improbable coincidences is very common and that it happens whenever, for example, a rock hits a window with more momentum than was necessary to make it break.) Now, as far as I can see, none of these other cases support an argument for disjunctive causes—at least not an argument of the type I have put forth here.

In sum, the Flipper and Reconnecter case differs from typical overdetermination cases in that it does not involve two causal paths to an outcome both of which go to completion, and which go to completion at exactly the same time. So, again, even if it is a case of overdetermination, it is an atypical example, and as such there are interesting lessons to be learned from it.[16]

III

Suppose that we are persuaded by the argument for disjunctive causes. What consequences does this have? In what follows I discuss three potential implications.

First, an obvious implication concerns the debate about the "causal relata." What kinds of things can enter in causal relations? If we allow for disjunctive causes, then, unless we are prepared to accept the existence of disjunctive events, we should be open to the possibility that other things besides events can enter in causal relations—probably facts. (This might not be a surprise, if we already believed that, say, omissions can enter in causal relations.)

A second, less obvious consequence concerns the concept of moral responsibility and its relation to causation. Imagine a scenario where we would want to blame Flipper and/or Reconnecter for the death of the person on the tracks. Imagine, for instance, that both Flipper and Reconnecter had reasons to refrain from doing what they did (for instance, because they did not know what the other would do, and thus they should have tried to prevent the death from happening by not doing their part). In that case the argument for disjunctive causes would have us say that, although each of them bears some moral responsibility for the outcome, neither causes it (more precisely, neither causes the outcome individually, although a disjunctive fact involving them does). If so, we should probably revisit some common beliefs about the relationship between moral responsibility and causation. For, according to the received view of their relationship, being morally responsible for an outcome requires, at the very least, causing it. In contrast, the argument for disjunctive causes suggests that it is possible to be responsible for something without being a cause in the standard sense.[17]

---

[16] Someone might think that the Flipper and Reconnecter case resembles a "preemptive prevention" case more than an overdetermination case (for discussion of preemptive prevention cases, see Michael McDermott, "Redundant Causation," *British Journal for the Philosophy of Science*, XLVI (1995): 523–44; and John Collins, "Preemptive Prevention," this JOURNAL, XCVII, 4 (April 2000): 223–34). Suppose that a catcher catches a ball that was flying towards a window; however, a second catcher was standing between the first catcher and the window and would have caught the ball otherwise. What caused the window to remain intact? Clearly, the second catcher did not do anything. But, given the presence of the second catcher, the window was in no danger of breaking, so maybe the first catcher was not a cause either (this intuition is reinforced if we substitute a solid wall for the second catcher). So maybe the cause was a disjunctive fact involving the two catchers? (The thought being that, if neither catcher had been present, the window would have shattered.) There are certainly similarities between preemptive prevention cases and the Flipper and Reconnecter case. But there are also important differences. One main difference is that preemptive prevention cases involve absences along the causal chain—for example, the ball's not hitting the wall—whereas the Flipper and Reconnecter case does not. This makes the Flipper and Reconnecter case an "ordinary" case (as defined in my introduction).

Finally, it is worth examining the bearing that the argument for disjunctive causes has, if any, for an argument for the autonomy of the mental from the physical. Roughly, the argument for the autonomy of the mental goes as follows.[18] Presumably, the mental is "multiply realizable" by the physical (this is to say, at least some mental events have different possible physical "realizers," or supervenience bases). But then this means that, for at least some mental events, there are no single physical events that could be identified with them across the board. So, unless we are prepared to identify mental events with disjunctive events (the disjunctions of all the different possible realizers of those mental events), we should probably conclude that mental events are not identical to physical events. Moreover, the argument would continue, the option of identifying mental events with disjunctions of physical realizers is problematic. Here the reasons offered might diverge or multiply. One reason could be that there are no disjunctive events. A related reason could be that there might be an infinite number of physical realizers for a given mental event, but that there are no infinite disjunctions, or events corresponding to infinite disjunctions. Another type of reason could be that, even if there were disjunctive events, or disjunctive events of the relevant type, they would not be able to enter in causal relations since there are no disjunctive causes (or that, even if there were disjunctive events, they could not occur in laws since generalizations about disjunctions are not sufficiently law-like). Note that this argument is likely to extend to subvenient/supervenient events in general; the case of the physical and the mental is simply an instance of that more general argument, but it is helpful to focus on just one instance.

---

[17] Typically we think that an agent causally contributes to an outcome just in case one of the agent's individual actions or omissions causally contributes to the outcome. In "How to Be Responsible for Something without Causing It" (*Philosophical Perspectives*, XVIII (2004): 315–36), I lay out the received view about the relationship between moral responsibility and causation, and I argue against it on the basis of an independent example. I then argue that moral claims are still dependent on causal claims, but not in the way specified by the received view.

[18] See, for example, Jerry Fodor, "Special Sciences," *Synthese*, XXVIII (1974): 97–115.

The question that arises, then, is: Can the argument for disjunctive causes be used to counter the argument for the independence of the mental from the physical? In particular, could it offer reasons to identify mental events with disjunctions of physical realizers on the grounds that those disjunctions could be causally efficacious, after all? In this case, my answer is negative: the type of scenario on which the argument for disjunctive causes was based is sufficiently different from the case of mental/physical events to not warrant drawing such an implication. In particular, given those differences, even if the argument for disjunctive causes succeeds in showing that there are *some* disjunctive causes, it has no tendency to show that disjunctions of *physical realizers* are causes. Thus it does not serve the purpose of defending the identification of mental events with disjunctions of physical realizers.

Let me explain. A main difference between the case of Flipper and Reconnecter and the case of a mental event and its physical realizers is the following. In the case of Flipper and Reconnecter, as I have pointed out, it is crucial that both actions were actual. If Flipper had acted without Reconnecter doing his part, or if Reconnecter had acted without Flipper doing his part, the outcome would not have been disjunctively caused. This is so because a main feature of the Flipper and Reconnecter case, and one *in virtue of which* I argued that we should say that the outcome was disjunctively caused, is that the two actions are "mutually neutralizing" (that is, they deprive each other of causal efficacy). But, for them to neutralize each other, they must both obtain. In other words, the argument for disjunctive causes that I have presented here crucially relies on a type of scenario where both (all) of the disjuncts are present. For each disjunct is needed to neutralize the causal powers of the other disjuncts (in particular, if only one of the disjuncts were present, then that disjunct would be a cause and the disjunction itself would not). As a result, the argument offers no reason to believe that causes can be disjunctive when only one disjunct is present.

Now, it is likely that, out of the potential physical realizers of a mental event, at most one of them will be actual at any given time. At least this is the way in which people normally think of the different possible physical realizers: as mutually exclusive physical states, or as states that are likely to be realized by organisms of different types (say, humans and Martians). This makes the case of mental events and their physical realizers importantly different from the Flipper and Reconnecter case. In virtue of this fact, the argument for disjunctive causes is not likely to have significant implications for the case of the mental and the physical.[19]

IV

I have argued that the initial presumption against disjunctive causes is misguided, and that there is a persuasive argument for disjunctive causes. I have also discussed some potential implications of that argument. Whereas I noted that it has important implications for the debate over the causal relata and for the relationship between causation and moral responsibility, we should not expect it to have significant consequences for an argument that the mental is autonomous from the physical.

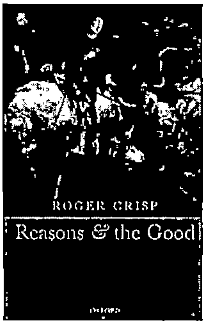
CAROLINA SARTORIO

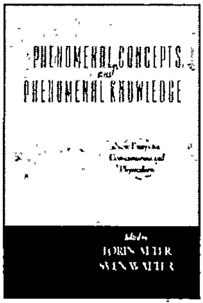
University of Wisconsin-Madison

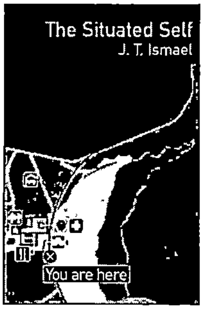
---

[19] In this sense, the case of the physical and the mental is more similar to the case of omissions. For, if omissions were disjunctive entities, they would be disjunctions of mutually exclusive disjuncts (for example, my failure to do *A* would be the disjunction of all the things that I could have done instead of doing *A*).

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Reasons and the Good
ROGER CRISP
Roger Crisp offers original and sometimes surprising answers to some of the deepest questions in philosophy, including: What reasons do we have for acting in one way or another? Are there moral reasons? What are reasons anyway? How can we know about them? What makes for a good human life? How should we weigh the well-being of others against our own?
October 2006 $45.00

Phenomenal Concepts and Phenomenal Knowledge
New Essays on Consciousness and Physicalism
TORIN ALTER and SVEN WALTER
What is the nature of consciousness? How is consciousness related to brain processes? This volume collects thirteen new papers on these topics: twelve by leading and respected philosophers and one by a leading color-vision scientist.
(Philosophy of Mind Series)
October 2006 $74.00

The Situated Self
J. T. ISMAEL
J. T. Ismael's monograph is an ambitious contribution to metaphysics and the philosophy of language and mind. She tackles a philosophical question whose origin goes back to Descartes: What am I? The self is not a mere thing among things—but if so, what is it, and what is its relationship to the world?
October 2006 $65.00

Structures of Agency
Essays
MICHAEL E. BRATMAN
This is a collection of published and unpublished essays by Michael E. Bratman, revolving around his "planning theory of intention and agency." Bratman's primary concern is with what he calls "strong" forms of human agency—including forms of human agency that are the target of our talk about self-determination, self-government, and autonomy.
October 2006 paper $35.00 cloth $99 00

Women's Liberation and the Sublime
Feminism, Postmodernism, Environment
BONNIE MANN
This book is a passionate report on the state of feminist thinking and practice after the linguistic turn. A critical assessment of masculinist notions of the sublime in modern and postmodern accounts grounds the author's positive and constructive recuperation of sublime experience in a feminist voice.
(Studies in Feminist Philosophy)
September 2006 paper $24.95 cloth $74.00

CAMBRIDGE

BEST IN SCHOLARSHIP

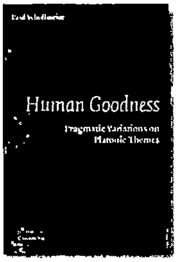
Nietzsche: 'On the Genealogy of Morality' and Other Writings
Second Edition
Edited by Keith Ansell-Pearson, Translated by Carol Diethe and Friedrich Nietzsche
Cambridge Texts in the History of Political Thought
$50.00: Hb: 0-521-87123-9: 216 pp.
$16.99: Pb: 0-521-69163-X

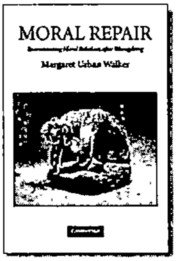
Kant's 'Critique of Pure Reason'
An Introduction
Jill Vance Buroker
Cambridge Introductions to Key Philosophical Texts
$75.00: Hb: 0-521-85315-X: 336 pp.
$24.99: Pb: 0-521-61825-8

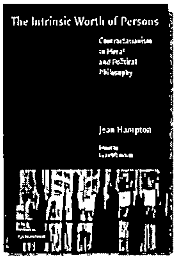
The Worlds of Renaissance Melancholy
Robert Burton in Context
Angus Gowland
Ideas in Context
$90.00: Hb: 0-521-86768-1: 356 pp.

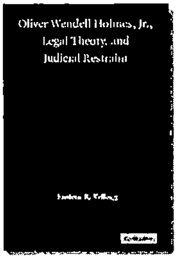
The Intrinsic Worth of Persons
Contractarianism in Moral and Political Philosophy
Jean Hampton
Edited by David Farnham
$70.00: Hb: 0-521-85686-8: 240 pp.
$26.99: Pb: 0-521-67325-9

Oliver Wendell Holmes, Jr., Legal Theory, and Judicial Restraint
Frederic R. Kellogg
$70.00: Hb: 0-521-86650-2: 208 pp.









The Cambridge Companion to Early Modern Philosophy
Edited by Donald Rutherford
Cambridge Companions to Philosophy
$75.00: Hb: 0-521-82242-4: 438 pp.
$29.99: Pb: 0-521-52962-X

Human Goodness
Pragmatic Variations on Platonic Themes
Paul Schollmeier
$80.00: Hb: 0-521-86384-8: 320 pp.

Moral Repair
Reconstructing Moral Relations after Wrongdoing
Margaret Urban Walker
$70.00: Hb: 0-521-81088-4: 264 pp.
$27.99: Pb: 0-521-00925-1

Fundamentals of Critical Argumentation
Douglas Walton
Critical Reasoning and Argumentation
$75.00: Hb: 0-521-82319-6: 360 pp.
$29.99: Pb: 0-521-53020-2

A Common Law Theory of Judicial Review
The Living Tree
W. J. Waluchow
Cambridge Studies in Philosophy and Law
$80.00: Hb: 0-521-86476-3: 296 pp.

The Cambridge Companion to Tocqueville
Edited by Cheryl B. Welch
Cambridge Companions to Philosophy
$80.00: Hb: 0-521-84064-3: 464 pp.
$29.99: Pb: 0-521-54996-5

Prices subject to change.

www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

New From The Routledge Philosophers Series

ARISTOTLE

Christopher Shields, Oxford University

"An impressive and first-rate overview of Aristotle's philosophy. I can't think of a better introduction."

—Richard Kraut, Northwestern University

March 2007: 256pp
Pb: 0-415-28332-9: $22.95

DARWIN

Tim Lewens, Cambridge University

"This book is a true gem. Although modestly billed as a '...philosophical introduction' to Darwin's thought, it's actually far more. With engaging style, Lewens deftly interweaves intellectual history and state-of-the-art evolutionary theory to illuminate current debates over ... 'intelligent design,' evolutionary psychology, evolutionary epistemology, ethics, sex differences, and the nature of science itself. Lively, timely, insightful, and informative, it's a terrific read."

—David Buller, Northern Illinois University

November 2006: 256pp
Pb: 0-415-34638-X: $22.95

HUSSERL

David Woodruff Smith, University of California, Irvine

"This is a first-rate volume...The text is rich with historical background and comment. But it is the clarity and depth of understanding with which David Woodruff Smith explains the material that is so overwhelmingly present. An excellent work."

—Gayle Ormiston, Kent State University

November 2006: 416pp
Pb: 0-415-28975-0: $22.95

KANT

Paul Guyer, University of Pennsylvania

"Kant is an absolutely first-rate general introduction to Kant's Critical Philosophy. Paul Guyer's interpretations are extremely well-supported, carefully and crisply argued, and highly insightful."

—Robert Hanna, University of Colorado

June 2006: 456pp
Pb: 0-415-28336-1: $22.95

RAWLS

Samuel Freeman, University of Pennsylvania

"Sympathetic, comprehensive, thorough, and accessible."

—Leif Wenar, University of Sheffield, UK

December 2006: 416pp
Pb: 0-415-30109-2: $22.95

Routledge
Taylor & Francis Group

1.800.634.7064 ☐ WWW.ROUTLEDGE.COM

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details.

THE JOURNAL OF PHILOSOPHY

VOLUME CIII, NUMBER 11
NOVEMBER 2006

page		
539	*The Metaphysics of Mental Causation*	Cynthia Macdonald and Graham Macdonald
	COMMENTS AND CRITICISM	
577	*Combination, Convention, and Possibility*	David B. Martens
	BOOK REVIEWS	
587	Isaac Levi: *Mild Contraction: Evaluating Loss of Informational Value Due to Loss of Belief*	André Fuhrmann
592	NEW BOOKS: ANTHOLOGIES	

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Libraries and Institutions (print and online)	$180.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.

POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2006 by the Journal of Philosophy, Inc.

ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CIII, NO. 11, NOVEMBER 2006

14.5.07
ALIPORE
CAMPUS

THE METAPHYSICS OF MENTAL CAUSATION*

A debate has been raging in the philosophy of mind for at least the past two decades. It concerns whether the mental can make a causal difference to the world. Suppose that I am reading the newspaper and it is getting dark. I switch on the light, and continue with my reading. One explanation of why my switching on of the light occurred is that a desiring with a particular content (that I continue reading), a noticing with a particular content (that it is getting dark), and a believing with a particular content (that by switching on the light I could continue reading) occurred in me, and these events caused my switching on of the light. This explanation works by citing the intentional contents of mental phenomena as causes of that action. It is because the intentional causes have the contents that they do, and because those contents play a causal role in bringing about my action, that my action is causally explained.

We think that a sufficient causal explanation can be given for my action solely in terms that make reference to its intentional causes and their intentional properties. However we have every reason to believe that there is a different causal story, a physical one, which can be told about my switching on of the light. This one will explain the physical movements that constitute my switching on of the light in terms that speak only of the physical events in my body and their properties. And we think that this story will be independent of the intentional one.

Assuming that intentional content does not reduce to physical properties of physical phenomena, we are faced with a problem. For

---

*We would like to thank John Carroll, Tim Elder, John Heil, Lawrence Lombard, Brian McLaughlin, Michael Pendlebury, David Robb, and Alan Weir for comments on an earlier draft of this paper.

0022-362X/06/0311/539–76

I act by moving my body. Indeed, my acting apparently just is my moving of my body as a (causal) result of certain of my mental states, certain beliefs and desires.[1] In this case, my switching on of the light just is my finger's flipping of the switch. So it looks like there is just one piece of behavior that needs explaining when I switch on the light, namely, the physical movements of my finger that constitute that action. We thus appear to have two competing causal-explanatory stories that we can tell about a single piece of behavior. One way of causally explaining the movements of my body that constitute my switching on of the light will mention only the mental properties of their intentional causes. Another way will mention only the physical properties of those causes.

As this way of developing the problem clearly illustrates, the difficulty cannot be resolved simply by maintaining that both stories allude to the same complex of internal physical events and processes as causes of the behavior to be explained. True though this may be, the problem has to do, not merely with the events that are invoked in the explanations, but with their properties. As long as the mental properties alluded to in the intentional causal explanation of the action are viewed as irreducible to, and so distinct from, the physical properties of those physical events, the problem of explanatory competition remains.[2] The explanations compete because each purports to causally explain the resulting behavior in terms of the causally relevant properties of its cause—the properties in virtue of which that cause brings about the effects it does—and each purports to be, at least in principle, a complete and independent explanation of that behavior.[3]

Let us call the problem in question here the *qua* problem—the problem that mental events that are physical events appear to be causally efficacious in the physical world *qua* physical but not *qua* mental.[4] The problem here is not one of causal efficacy of events, but rather, one of what we shall call *causal relevance* of properties.[5] And let us call the position on the relation between the mental and physical that gives rise to this problem *minimal physicalism*. Minimal physicalism is the view that although each mental event or phenomenon is a physical event or phenomenon, mental properties are wholly distinct from physical ones.

---

[1] On the view being articulated here, my moving of my body just is my causing of my body's moving, and is identical with my switching on of the light (which is my causing of the light to come on).

[2] The notion of irreducibility invoked here is specific to accounts of reduction that require of a reduction that it effect a property-identity uniting the two domains (as in the account offered by Robert Causey, *The Unity of Science* (Boston: Reidel, 1977); Causey's "attributes" are our properties). These have been called "conservative" reductions, the reduction preserving the reduced property. Property nonidentity and irreducibility go together on these accounts. Eliminative, or replacement, accounts of reduction are not relevant here, as they would ensure the irrelevance of the replaced properties.

[3] As this makes clear, we assume the standard view that causal explanations cite causally relevant properties of the causes of the effects being explained. See, for example, Jaegwon Kim, "Explanatory Realism, Causal Realism, and Explanatory Exclusion," in Peter A. French et al., eds., *Midwest Studies in Philosophy,* Volume 12 (Notre Dame: University Press, 1987), pp. 225–39. The idea of a "complete and independent" explanation derives from Kim's discussion of what he calls "The Principle of Explanatory Exclusion," which we discuss in the text (pp. 544–45). (See, for example, his "Mechanism, Purpose, and Explanatory Exclusion," in James E. Tomberlin, ed., *Philosophical Perspectives,* Volume 3 (Atascadero, CA: Ridgeview, 1989), pp. 77–108.)

It is worth noting that minimal physicalism can generate problems not only for intentional explanations, but also for any other kinds of explanations, say biological ones, which make essential use of nonphysical properties in the proffered explanations. For any such set of explanations the question will arise: in virtue of which property/ies does the effect occur? The threat will be the same as it is in the mental case, that the nonphysical properties cited in the explanation do not produce the effect *qua* those properties. We are here concentrating on the mental, on intentional explanations and intentional properties, because the threat has received its sharpest formulation, and has been most widely discussed, for this case. It has been argued that the problem does not have the general application we think it has, but whether it does or not will not be relevant to our proposed solution.[6] That solution will generalize to other relevantly similar sets of explanations, if there are any, as will become evident in section v.

---

[4] The debate about causal relevance has roots in the seminal work of Donald Davidson and the position he calls "anomalous monism." See his "Mental Events," in *Essays on Actions and Events* (New York: Oxford, 1980), pp. 207–24, and "Psychology as Philosophy," along with his "Comments and Replies," pp. 229–39 and 239–44. Davidson argues that psychophysical event interactions are only covered by laws couched in physical terms, so it appears that mental causes only count as such in virtue of their physical properties. For a defense of causal relevance of mental properties within the Davidsonian framework that differs from ours, see Terence Horgan, "Mental Quasation," in Tomberlin, ed., *Philosophical Perspectives*, Volume 3, pp. 47–76.

[5] In earlier work, we distinguished between causal efficacy of events and causal relevance of properties, taking the former to concern events in extension, and the latter to concern *both* the *qua* problem and the explanatory potential of properties of events (see "Mental Causes and Explanation of Action," *Philosophical Quarterly*, XXXVI (1986): 145–58, reprinted in Leslie Stevenson, Roger Squires, and John Haldane, eds., *Mind, Causation, and Action* (New York: Blackwell, 1986), pp. 35–48, and "How to Be Psychologically Relevant," in Macdonald and Macdonald, eds., *Philosophy of Psychology: Debates on Psychological Explanation* (Cambridge: Blackwell, 1995), pp. 60–77). We now think that the problem that goes under the name of 'causal relevance' concerns the problem of how mental events can bring about effects in virtue of their mental properties (the *qua* problem, or the problem of "too little" causal relevance), and that this is an issue concerning the causal powers of mental properties. This issue is distinct from, though intimately related to, that of explanatory relevance. We also think that much of the criticism directed toward nonreductive monism, and in particular, the "too much" causal relevance charge, conflates issues concerning causal efficacy of events with ones concerning causal relevance of properties. For an example of someone who moves from talk of one to talk of the other, see Stephen Yablo's "Mental Causation," *The Philosophical Review*, CI (1992): 245–80, and quoted material in note 14. Davidson ("Thinking Causes," in John Heil and Alfred Mele, eds., *Mental Causation* (New York: Oxford, 1993), pp. 1–17) argues that, because the causal relation holds between events in extension, it makes no sense to speak of an event's causing another in virtue of one or another of its properties. However, Brian McLaughlin demonstrates how one can make clear sense of such talk even given the claim about causation. See his "On Davidson's Response to the Charge of Epiphenomenalism," in Heil and Meles, eds., *Mental Causation*, pp. 27–40.

Minimal physicalism, and more specifically, nonreductive monism, has been discarded by many as a viable position on the mind-body relation on the grounds that it gives rise to, but cannot resolve, the problem of the causal relevance of the mental.[7] In what follows, we want to defend a strategy—one that we argued for some time ago—for solving this problem. Specifically, we want to defend the view that nonreductive monism can resolve the problem if it is combined with a certain metaphysical view of the nature of events, the property exemplification view. This defense contrasts with another, increasingly popular, strategy for dealing with the problem, one that involves appeal to the view that mental events and other phenomena are, or involve, tropes. Those who endorse it think not only that the trope strategy can solve the problem of the causal relevance of the mental consistently with minimal physicalism, but also that appeal to other metaphysical views of events and properties (specifically, the property exemplification account) cannot do the job as well (if at all), so that appeal to trope theory has the explanatory edge over appeal to other metaphysical theories of events and properties.[8]

Since the view that nonreductive monism is a philosophically bankrupt position unless (and, perhaps, even if) it is combined with a trope conception of events and properties has become entrenched in the literature, we think it is time to set the record straight. In what follows we argue both against those who think that appeal to trope theory can rescue causal relevance of the mental for nonreductive monism, and against those who think that nonreductive monism is not a viable position on the mind-body relation because it cannot resolve the problem of the causal relevance of mental properties. The charge of causal irrelevance against nonreductive monism has two aspects: one specifically concerns the *qua* problem—the problem of how mental events can bring about effects in virtue of their mental properties (the "too little" causal relevance problem)—and the other concerns the complaint that the position cannot solve the *qua* problem in a way that does not lead to there being "too much" causal relevance for mental properties.

---

[6] Kim has argued (in *Mind in a Physical World* (Cambridge: MIT, 1998)) that the problem does not generalize; we think it does, for the reasons put forward by Thomas Bontly in "The Supervenience Argument Generalizes," *Philosophical Studies*, CIX (2002): 75–96.

[7] See, for example, Louise Antony, "Anomalous Monism and the Problem of Explanatory Force," *The Philosophical Review*, XCVIII (1989): 153–87; Fred Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge: MIT, 1988), and "Reasons and Causes," in Tomberlin, ed., *Philosophical Perspectives*, Volume 3, pp. 1–15; Ted Honderich, "The Argument for Anomalous Monism," *Analysis*, XLII (1982): 59–64; Kim, "Epiphenomenal and Supervenient Causation," in his *Supervenience and Mind: Selected Philosophical Essays* (New York: Cambridge, 1993), pp. 92–108, and *Mind in a Physical World*; and Yablo, "Mental Causation."

[8] See especially Douglas Ehring, "Mental Causation, Determinables and Property Instances," *Noûs*, XXX (1996): 461–80, *Causation and Persistence: A Theory of Causation* (New York: Oxford, 1997), and "Tropeless in Seattle: The Cure for Insomnia," *Analysis*, LIX (1999): 19–24; and David Robb, "The Properties of Mental Causation," *Philosophical Quarterly*, XLVII (1997): 178–95, and "Reply to Noordhof on Mental Causation," *The Philosophical Quarterly*, LI (2001): 90–94.

We begin, in section I below, by briefly outlining the position, presuming the distinction, crucial to it, between causal efficacy of events and the causal relevance of properties of events, and explain how the *qua* problem arises. In sections II and III, we clear the ground for our defense by exploring the trope-physicalist solution to the *qua* problem. We argue that that appeal to trope theory does not effectively deal with it. In section IV we outline the version of the property exemplification account that we favor, and explain why the claim, made by those who appeal to the trope strategy, that appeal to this alternative account cannot help to solve the problem of the causal relevance of mental properties is false. Crucially, we argue that the property exemplification account has the resources to explain how mental properties of events *could* be causally relevant to the effects their instancings bring about without overdetermining them, and without competing for causal relevance with the physical properties of those events. Finally, in section V, we defend our own strategy for dealing with the problem of the causal relevance of the mental, making use of our version of the property exemplification account.

I. FOUR THESES

How exactly does the charge that mental properties are causally irrelevant arise? Suppose that the *minimal physicalism* (MP), described above, is true, each mental event or phenomenon being identical with a physical event or phenomenon. What makes the physicalism minimal is that the mental properties of mental/physical events, properties such as *being pain*, or *being a coming to believe that it is getting dark*, are assumed to be distinct from, and irreducible to, the physical properties of those events, such as *being C-fibers' firing*.[9] Then one claim that is involved in the charge that mental properties are causally irrelevant is this:

---

[9] Again, we are assuming here that causal claims and explanations not only cite events, but also cite properties of events in virtue of which they are causally related, a view that is held by Kim ("Explanatory Exclusion and the Problem of Mental Causation," in

(PCR) Physical properties of physical events are causally relevant to the physical effects those events bring about.

Let us call the thesis embodied in PCR the *Principle of the Causal Relevance of Physical Properties.* What it says is that physical properties of physical events are such that their instances are causally effective in bringing about physical effects of those events. Analogously, let us call the thesis embodied in the following claim:

(MCR) Mental properties of physical events are causally relevant to some of the mental and physical effects those events bring about.

the *Principle of the Causal Relevance of Mental Properties.* MCR is the denial of the claim that only physical properties are causally relevant, and so is the thesis whose defense is at issue.

PCR and MCR, taken together, generate causal tension when conjoined with a third claim, which we can call *Exclusion*:[10]

(EXCL) If a property, *P*, of a cause, *c*, is causally sufficient for an effect, *e*, then no other property, *Q*, distinct from and independent of *P*, is causally relevant for *e*,

where *P*'s being causally sufficient for *e* means that an instance of *P* in an event *c* is causally sufficient for *e*. It follows from principle EXCL that mental properties and physical properties of physical events cannot both be completely and independently causally relevant to a single effect.

A brief comment on the origin and purport of EXCL is in order. Its origin is Kim's Principle of Explanatory Exclusion, which states, "No event can be given more than one *complete* and *independent* explanation."[11] Kim here leaves the notions *complete* and *independent* undefined, leaving it to our intuitions on particular cases to tease out their meaning, but it is fairly clear what restrictions need to be placed on any relation that can exist between two putatively independent *causes*, c_1 and c_2, of the same event, e_1: they cannot be identical; they cannot be partial causes, constituents of a larger cause; they cannot be different links in the causal chain leading to e_1; neither cause can be a "part" of the other cause.

Enrique Villaneuva, ed., *Information, Semantics, and Epistemology* (Cambridge: Blackwell, 1990), pp. 36–56), and Dretske ("Reasons and Causes"), amongst others (see note 7).

[10] Barring cases of overdetermination.

[11] Kim, "Mechanism, Purpose, and Explanatory Exclusion," p. 79.

Explanatory completeness is to be understood in terms of sufficient causation, so if c_1 is causally sufficient for e_1 there will be a description of c_1 that *will* yield an independent and complete explanation of e_1 (since there will be no c_2 such that c_1 and c_2 are independent causes of e).[12] The notion of "a description" of an event is too vague, however, as events will have many descriptions that do not yield any causal explanations of their effects, so a more precise account must limit the descriptions to those that mention causally relevant properties. This version of EXCL is a descendant of Kim's, and is similar to Yablo's property version of a principle he labels *exclusion*, which reads: "If a property *X* is causally sufficient for an event *y*, then no property *X** distinct from *X* is causally relevant to *y*."[13] (Again, the notion of causal sufficiency is undefined, except to say that it could mean sufficient-in-the-circumstances or sufficient absolutely.) The present version is more explicit on what it means to say a property is causally sufficient, since this is necessary to avoid the conflation of causal efficacy with causal relevance.[14]

Given principles PCR, MCR, and EXCL, it might look as though a situation in which there are two causal explanations of a single piece of behavior such as my switching on of the light—one that cites only the mental properties of its mental/physical cause, and one that cites only the physical properties of that same cause—is a situation in which we must choose between them.

However, in order to generate the charge that mental properties are causally irrelevant (that is, that MCR is false), something further is needed. What is needed is commitment to a view about physical causation and explanation. This view is that the physical domain is closed under causation and causal explanation. (This is sometimes referred to as *Closure*.[15]) Closure states:

---

[12] Thanks here to John Carroll.

[13] This derived from Yablo ("Mental Causation," p. 247), in line with his instructions (in his note 5) for translating the event-version into the above property-version of *exclusion*.

[14] Yablo claims that "any credible reconstruction of the exclusion principle must respect the truism that determinates do not contend with their determinables for causal influence," p. 259. We agree, but he questions our rationale, using the example of Ella's singing at 80 db, where her singing at more than 70 db shatters a glass. His claim is that our attempt to explain why the property of singing at 80 db does not compete for causal influence with the property of singing at under 90 db either gives a result too strong ("too much" causal relevance) or is unsatisfactory because it makes properties "efficacious not absolutely, but only relative to some specified effect." This is a clear example where issues concerning causal efficacy of events are conflated with issues concerning causal relevance of properties. In Yablo's example, both properties are ones whose exemplifyings (events) are causally efficacious, but this does not show that both the property of being a singing at 80 db and the property of being a singing at under 90 db, properties of those events in virtue of which they are causally efficacious, are causally relevant, nor that they "compete for causal influence," since causal efficacy of events is only a necessary condition on causal relevance of properties of those events. As this last point suggests, we also think that Yablo's discussion conflates properties whose exemplifyings (by persons) are (that is, are identical with) events with properties *of* events. The issue of causal relevance, the *qua* problem, concerns properties *of* events.

> (CLOS) If a physical event or phenomenon has any cause, it has a sufficient physical cause, whose physical properties are causally sufficient for its effect.[16]

Assuming that causal explanations are at least cause- and effect-citing, it follows from Closure that if a physical event has any causal explanation, it has one that alludes only to its causally sufficient physical properties.

How does CLOS work, together with PCR, and EXCL just mentioned, to generate the charge that MCR is false? Well, we know from MCR, PCR, and EXCL that if we have two properties (one mental and one physical) of a cause, each of which can be cited in two causal explanations (one mental and one physical) of a single explanandum, we cannot (by EXCL) accept both unless neither of those properties, taken alone, is causally sufficient. What CLOS entails is that there is some causally sufficient (physical) property.

In short, the combination of PCR, EXCL, CLOS, and MCR appears to be inconsistent. If, then, we accept the former three principles, it seems that we must reject MCR and with it the view that mental properties of events are causally relevant to the effects those events bring about. But we want *both* to accept the former three principles and to accept MCR.

II. TROPES

The trope-physicalist claim is that appealing to trope theory offers a way of avoiding the charge that mental properties are causally irrelevant while maintaining PCR, EXCL, and CLOS, and, of course, the requirement on minimal physicalism that mental and physical properties are irreducibly distinct. In order to assess this claim, we need to know something about the theory of tropes on which it is based.

Consider a particular cardinal bird sitting on the branch of the tree just outside my window. It has certain properties: it is red, it has a particular size, shape, position, and so on. According to trope theory, these properties are particular properties of the bird. They are as particular as the bird itself; they are not particular instances of a property that can also be instantiated at the same or distinct times in other places or objects.

---

[15] See, for example, Robb, "The Properties of Mental Causation," and citations therein.

[16] In recent work Kim has given this formulation of Closure: "If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*" ("Blocking Causal Drainage and Other Maintenance Chores with Mental Causation," *Philosophy and Phenomenological Research*, LXVII, 1 (July 2003): 151–76, here p. 158). Our formulation strengthens his to emphasize sufficient causation, which, together with EXCL, links closure to causal *relevance* for reasons that will become clear. Robb (in correspondence) has pointed out that PCR is strictly redundant, given EXCL and CLOS. We retain it just to emphasize the role that relevance plays in this debate.

More precisely, in standard philosophical usage a property is construed as a universal, and an instance of a property is not a trope of that universal, but a thing that has (instantiates, exemplifies) that property. So a cardinal is an instance of the property of being red. For trope-theorists, however, the so-called particular redness of the cardinal, a trope, is not an instance of redness; it is a trope of redness. Just as the cardinal cannot be in more than one place at any given time—just as it is located all at once, wholly and completely, in the place it occupies at any given time—so too, its redness, its size and its shape cannot be in more than one place at any given time.

Tropes have been called by many names, one common one being "abstract particular." Keith Campbell tells us that resistance to the idea of a trope is based on a conflation of one pair of terms or concepts, that of universal and particular, with another, that of abstract and concrete.[17] This conflation is responsible for the belief that since to be universal just is to be abstract, to be particular just is to be concrete. On this basis, the possibility that there should exist things that are both abstract and particular is ruled out.

However, Campbell goes on to say, what is universal is so because it is possible for it to be wholly and completely in more than one spatial position at any given time. In contrast, what is particular can only be wholly and completely in one spatial position at a given time. And this is different from the contrast between abstract and concrete. What is abstract is "what is got before the mind by an act of abstraction." It is what comes to be known by attending to some part, or aspect, of "what is presented."[18] In contrast, what is concrete is grasped by attending to all of what is presented, not by attending to some part or aspect of it. No special act of selective attending is involved in our grasp of concrete things.

---

[17] Campbell, "The Metaphysic of Abstract Particulars," in French et al., eds., *Midwest Studies in Philosophy,* Volume 6 (Notre Dame: University Press, 1981), pp. 477–88; and *Abstract Particulars* (Cambridge: Blackwell, 1990).

[18] Campbell, *Abstract Particulars,* p. 48. This sense of 'abstract' is different from another, and more common understanding of the term in metaphysics. According to it, entities of a given kind are abstract if they are such that many of them can be in the same place at the same time. It contrasts with the sense of 'concrete', where to be concrete is to exclude other things of the same kind occupying the same space at the same time.

Campbell's work is a revival of the earlier foundational work of Donald C. Williams.[19] Williams set forth a version of trope theory that is now regarded as the classic account.[20] This account will serve as the standard against which to evaluate the strategy of appealing to trope theory in order to resolve the problem of the causal relevance of mental properties.

According to the classic account, tropes are the foundations of all things, and so are the foundations of both individual objects and events, and properties or universals if there are any. Individual objects are bundles of compresent or concurrent tropes, tropes that are "bundled" together by relations of compresence. For example, this redness, this shape, this size, and this position (along with other tropes), related to one another by compresence relations, together "constitute" or comprise the cardinal sitting on the branch of the tree outside my window. And individual events are either tropes, or bundles of compresent tropes.[21]

Properties or universals, and relations are classes or sets of exactly resembling tropes. Individual tropes (for example, all the rednesses of all the red things) that are members of the class or set of exactly resembling tropes with which a given property is identical are understood to be particular trope-instances of that property. For a substance or event to instantiate a property (universal, relation) is for the intersection of the set of compresent tropes that constitute the individual substance or event and the set of tropes that is the property to be nonempty.

III. TROPE THEORY AND THE PROBLEM OF THE CAUSAL RELEVANCE OF MENTAL PROPERTIES

We are now in a somewhat better position to appreciate the basic commitments of trope theory and so of its prospects for resolving the conflict between MCR, PCR, EXCL, and CLOS, assuming minimal physicalism. The strategy appeals to a trope conception either of events, or of the mental and physical tropes that constitute events, in order to make sense of the claim that each mental event is identical with a physical event, and to a trope conception of properties in order to make sense of the claim that mental and physical properties are irreducibly distinct. According to it, to say that mental events are identical with physical events is to say that mental tropes and physical tropes are identical, where these tropes either are events, or constitute events and so are "parts" of them. If the latter, events are, like any other object, complexes of many distinct tropes. Either way, because mental tropes are identical with physical tropes, the claim is that there is no problem about the causal efficacy of mental events. On both understandings of the trope conception, psychophysical event identity is guaranteed by psychophysical trope identity. Given that events are the relata of the causal relation, psychophysical trope identity guarantees the causal efficacy of mental events.

---

[19] Williams, "The Elements of Being," *Review of Metaphysics*, VII (1953): 3–18, 171–92; cited from *Principles of Empirical Realism* (Springfield, MA: Charles C. Thomas 1966), pp. 74–109.

[20] John Bacon, *Universals and Property Instances: The Alphabet of Being* (Cambridge: Blackwell, 1995), and "Tropes," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Fall 2002 Edition on the Internet). Bacon refers to the classic account as "trope cluster theory."

[21] Robb (in correspondence) claims that he assumes neither of these views in "The Properties of Mental Causation," although he does in that paper say that the trope solution can be combined with a view of events as Kim-like structured particulars, and agrees that on this understanding of an event the causal efficacy/causal relevance distinction collapses on the trope strategy. In the paper, he takes events to be basic concrete Davidsonian ("unstructured") particulars that "instantiate" tropes. But how are we to understand the claim that events "instantiate" tropes? Robb (again in

What does the question of causal relevance of mental properties amount to on the trope conception, given minimal physicalism's commitment to the irreducibility of mental properties? It depends, since, on the trope conception, 'property' is ambiguous: it might mean either "trope" or "class of tropes." Depending on how it is interpreted, there will be a different way of understanding the question. Accordingly, there are two ways in which in the defense of MCR might be mounted.[22] We will begin by considering the "class of tropes" interpretation.

According to this conception, mental properties or types are sets or classes of exactly resembling mental tropes, and physical properties

---

correspondence) takes the instantiation relation to be a relation between a particular and a property, irrespective of whether the latter is understood as a universal or as a trope. Understood this way, tropes characterize events. For example, phenomenal events, such as pains, hurt, and so have phenomenal tropes, and intentional mental events, like thinkings, have contents, in that they are characterized by certain content-tropes. However, given the distinction between causal efficacy of events and causal relevance of properties, this way of understanding the relation between an event and its trope-properties makes, say, the hurt-trope of a pain *inefficacious* in that event's bringing about the effects it does, and similarly for the content-trope of a thinking. Quite apart from the implausibility of this view, if tropes of events are inefficacious it is hard to see how they could be causally relevant, since it is natural to suppose that a necessary condition on causal relevance of a property (universal or trope) is that its instantiation is causally efficacious. The more natural (and more plausible) reading of "instantiate" for the trope theorist, then, is one according to which an event instantiates a trope by being constituted by it.

[22] Ehring defends MCR by appeal to the class of tropes conception of mental properties, whereas Robb defends MCR by appeal to the tropes conception.

or types are classes of exactly resembling physical tropes. The strategy ensures the distinctness of the mental property-classes and the physical property-classes by viewing mental properties as higher-level properties of the events whose physical properties are lower-level ones, in something like the way in which, say, being soluble is a higher-level property of substances, such as salt, that have lower-level physical properties. That is to say, they are higher-level properties in virtue of the fact that although they are first-order properties of particular things like events and substances, they are possessed by those things in virtue of the possession by those things of other (first-level) properties. Thus, for example, being soluble is possessed by sugar cubes in virtue of those cubes' having the chemical constitution they have.[23]

Suppose that a necessary condition on causal relevance of a property or type is causal efficacy of its property instance or trope. Then, on this understanding of what it is to be a mental type or property, to say that mental tropes that are identical with physical tropes are causally efficacious in virtue of their mental properties is to say that the mental tropes that comprise mental/physical events fall into two classes, a mental one and a physical one. Being physical tropes, they fall into a class of exactly resembling physical tropes. This class, class A, is determined to be the class it is in virtue of resemblance relations that hold between its members. The very same tropes that fall into this class (as well as ones that fall into other classes of physical tropes), however, fall into another, higher-level class, class B. They do so in virtue of falling into the first-level classes that they do.

Class B is higher-level: it is a class of tropes whose members fall into it in virtue of falling into other classes of exactly resembling physical tropes. This class is a class of mental tropes, and it is a class whose members are physical tropes that are not exactly resembling in ways that make for their membership in any physical class. They are not exactly resembling because class B's tropes fall into distinct classes of exactly resembling physical tropes, such classes being distinct because the members of each such class do not exactly resemble the members of the others. Thus, for example, the mental property, being pain, is a class of tropes, whose members fall into it in virtue of falling into other, first-level classes, such as, for example, the class, being C-fibers' firing, the class, being H-fibers' firing, and so on. On the assumption that mental properties are variably realizable by physical ones, there may be an indefinite number of such first-level classes.

---

[23] Higher-level properties should not be confused with higher-order ones. Higher-order properties are properties of properties, not properties of the things that have them in virtue of their possession of other properties. It is common, especially in functionalist treatments in the philosophy of mind, to use 'higher-order' rather than 'higher-level' when talking about mental properties such as being pain, or dispositional properties like being soluble. But this is quite different from the contemporary logician's usage (though similar to Russell's and Ramsey's). In contemporary terms, 'being soluble', like 'being a number', is a first-order predicate and so stands for a first-order property because its instances are particulars. However both predicates might be classed as impredicative, that is, specificiable by phrases that include second-order quantification over all properties, including those properties themselves. Thus, 'λx(x is soluble)' might be specified by something like, '$\exists F(Fx$ & $\forall y(Fy$ & y is placed in a relevant liquid$\rightarrow y$ dissolves)', where we have a second-order quantifier, '$\exists F$', which ranges over all properties, including being soluble (just as the bound variable in '$\iota x(\forall y(x \; 1 \; y \rightarrow x$ is taller than y)' impredicatively specifies the tallest person. Thanks to Alan Weir for discussion on this point.

It is crucial here that the resemblance relations binding class B's members together are not the same resemblance relations as those that bind the members of each first-level physical class together. The resemblances that bind the members of the higher-level class together are themselves higher-level resemblances. Consider, again, the higher-level property, being soluble. Members of the class that, on the present version of the strategy, just is this property are exactly resembling in this way: each thing (substance) that has (/is partly constituted by) a member (trope) of that class is apt to dissolve if placed in a relevant liquid. The property itself is higher-level, because its members exactly resemble one another in this way in virtue of falling into other, chemical compound classes (for example, the *NaCl* class). Members of this lower-level class are chemical tropes. These very tropes, however, fall into higher-level classes in virtue of being members of a class of exactly resembling chemical tropes.

Suppose, now, that we define a given mental property, or class of tropes, in terms of its causal role. Its members are occupants of that causal role; they are what discharge it. Then the claim is that physical tropes, which have physical properties in that they are members of physical classes, are the occupants of the causal role definitive of the mental property or class. It follows that mental tropes are physical tropes. This view we shall call *trope functionalism,* since it is a functionalist account of the relation between the mental and the physical, applied to tropes.[24]

There are several advantages to this position. First, by maintaining that mental and physical tropes are identical, it ensures the causal efficacy of mental events.[25] Second, by maintaining that the tropes that fall into both physical and mental classes are physical tropes, it ensures the primacy of the physical, and with it the vindication of principles CLOS and PCR. Finally, by maintaining that mental properties are second-level classes of physical tropes that fall into first-level (physical) classes, it ensures the distinctness of mental and physical properties. Further, it seems capable of meeting the Exclusion requirement, since the causal roles by which mental properties are defined open up space for physical tropes to occupy. Given the distinction between role and occupant, there seems to be no real problem of exclusion to address.

---

[24] Robb, unlike Ehring, does not specifically advocate trope functionalism with respect to mental property-types, but advocates what he calls "trope monism." However, he distinguishes mental property-types and physical property-types, where types are taken to be universals or classes, and construes such types as classes or sets of resembling tropes. On this basis he claims, "if trope monism is true, a given mental type is a set of physical tropes. But multiple realizability entails that these physical tropes do

This leaves MCR and the charge of causal irrelevance of mental properties. But here the strategy founders. Either it is landed with a version of the *qua* problem, and so cannot handle the problem of the causal relevance of mental properties, or it avoids the *qua* problem, but only by sacrificing MCR altogether.

Let P_1, P_2, P_3, P_4 be physical property-classes, each being comprised by exactly resembling physical tropes, and let M be a mental property-class. Here we are talking about mental and physical property-types. Then the trope view is that M is a class of physical tropes (that are not exactly resembling): it is the class whose members just are the members of P_1, P_2, P_3, and P_4.

Suppose now that particular P_i in P_1 causes an effect e. We can ask, in virtue of which property-type does P_i cause e? Does it cause e in virtue of being a member of P_1 or in virtue of being a member of M? If as seems inevitable, the answer is: in virtue of being a member of P_1, then PCR is purchased but at the expense of the causal relevance of the property-type M. So the principle MCR is jettisoned. It is jetti-

---

not themselves resemble one another in the way that members of a physical type must: they will be wildly dissimilar physically. So the mental type is not itself a physical type (hence *Distinctness*), though of course it has many physical types as subsets: these just are the physical types that 'realize' the mental one" (Robb, "The Properties of Mental Causation," p. 188). This is consistent with a form of trope functionalism with regard to mental types. Ehring explicitly takes mental properties (property-types) to be a species of dispositional properties, takes dispositional properties to be a species of functional properties, and takes functional properties to be second-order properties as defined thus: "A second-order property is the property of having some property X or other that satisfies a certain specification H" ("Tropeless in Seattle," p. 19). So, both Ehring and Robb are trope monists, and both Ehring and Robb take mental property-types to be second-level classes of physical tropes.

[25] This is so irrespective of whether one endorses the view that it is events that are the causal relata or whether one endorses the view that it is tropes that are the causal relata. On a trope view, events are "bundles" of compresent tropes, so if mental events are causally efficacious, then mental tropes of mental events are causally efficacious.

soned because M is a higher-level property, whose causal relevance is purchased only by its members' (physical tropes) being physical (that is, being members of P_1).[26]

Suppose that the trope theorist responds by claiming that, since one and the same trope, P_i, is (that is, is identical with) an instance of both P_1 *and* M (that is, there is just one, physical trope here), if P_1 is causally relevant so too is M. Then another, different problem arises. Consider once again the higher-level property, being soluble, and one of its lower-level realizing properties, being NaCl. All of the tropes that are members of the being NaCl class are also members of the being soluble class. Nevertheless, we want to say that the properties are distinct, since the former is a first-level class whereas the latter is a second-level one.

The problem is not that the strategy cannot ensure the distinctness of first-level and second-level properties, since it can.[27] However, as the being soluble example makes plain, the distinction between higher-level and lower-level properties does not by itself ensure that those higher-level properties that things have in virtue of having lower-level physical properties are not themselves physical properties. For *being soluble* is a physical property! Applied to the psychophysical case, there is no reason to call the higher-level property/class "mental." In fact, we have every reason not to if all its members (that is, tropes) are physical.

Of course, it is open to a trope functionalist to respond by saying that the earmark of the mental is not having a functional role *per se*, but having a *particular* functional role, and that the functional role that binds all the being soluble tropes together just is not the right sort of functional role. But even if this response is acceptable, a more fundamental problem with the strategy remains, and that is that it simply does not follow from the fact that P_i is an instance of both P_1 and M_1 that if P_1 is causally relevant, so is M_1. P_i will be an instance of any number of property-types, at least some of which (such as the property-type that is some heterogeneous class whose members bear only some weird kind of resemblance to one another) are plainly not causally relevant. Given this, we really have no reason to think the M_1 is causally relevant. This is so even if the resemblance between the members of M_1 is not functional.[28]

---

[26] Robb ("Reply to Noordhof on Mental Causation") interprets Noordhof ("Do Tropes Resolve the Problem of Mental Causation?" *Philosophical Quarterly*, XLVIII (1998): 221–26) as objecting to his position for something like this reason. Robb's reply is that it makes no sense to speak of tropes being causally relevant in virtue of their properties, if properties are construed as types or classes, since classes are not the sort of things that can be causally relevant. So Robb, though not Ehring, does not endorse this version of the trope strategy. We develop it and our objection to it here partly because it does apply to Ehring's version, and partly because, as we shall see, it is not clear that Robb's attempt to distinguish causal efficacy of events from causal efficacy and causal relevance of properties at the level of tropes can succeed (see also note 21), in which case his strategy too fails to deal with the problem of causal relevance of mental properties.

[27] For, even if we consider all the possible occupants of the causal role definitive of a higher-level property, such as being soluble, there is still the distinction between the property specified impredicatively, that is, specifiable by phrases that include second-order quantification over all properties, including those properties themselves, as, say, '$\exists F(Fx \,\&\, \forall y(Fy \,\&\, y$ is placed in a relevant liquid $\rightarrow y$ dissolves)' and the things that truly instantiate F, that is, all assignments to x which satisfy it in virtue of falling into lower-level, physical classes.

So this version of the trope strategy by itself does not ensure the truth of the principle MCR. But there is a deeper problem for the trope-physicalist lurking here. To see this, we need to look at the second avenue available to the trope theorist for defending MCR mentioned earlier in connection with the two ways of understanding 'property' recognized by trope theory.

The trope theorist is not forced to defend MCR by construing mental properties as property-types, that is, classes of tropes. She can maintain instead the *qua* problem can only arise at the level of individual tropes, and insist that if one accepts that mental properties are higher-level classes of physical tropes whose members fall into them in virtue of falling into lower-level classes, then there is no problem about mental causation or causal relevance at that level.[29] The grounds given for this might be that the items doing the work are not the classes but their members (the tropes); and no member does anything in virtue of its class membership. Indeed, it might be claimed, it makes no sense to say that a given trope caused (or figured in a cause of) *e* in virtue of being in one class or another.

Pursuing this strategy seems to avoid both of the above charges. By rejecting the view that the problem of causal relevance of mental properties concerns property-types altogether, it apparently avoids the charge that the *qua* problem arises again at the level of property-types. But it also appears to avoid the charge that MCR is jettisoned because there is no reason to think that higher-level classes of physical tropes are anything other than physical properties. For, according to the present way of understanding the problem of causal relevance of properties, though it may be true that higher-level classes of physical tropes are themselves physical property-types, it is irrelevant to the question of whether mental properties are causally relevant, this question being one that concerns individual tropes, not classes of tropes.

---

[28] We are indebted here to John Carroll.

[29] This is the strategy that Heil and Robb pursue in "Mental Properties," *American Philosophical Quarterly*, XL (2003): 175–96. See also Heil's *From an Ontological Point of View* (New York: Oxford, 2003).

But it is not at all clear that the question concerns only tropes. According to the trope theorist, the trope strategy does not face the *qua* problem at the level of property-types, but this is not to say that there is no problem of causal relevance of properties, as distinct from causal efficacy, that arises at the level of tropes. According to the present strategy, when a given mental event, c, causes another event, e (where an event is an object's possessing a property at a time and where property = trope), there are two questions that can be asked, one about causal efficacy of events or event-tropes and another about causal relevance of properties. We can ask what that event c causes, and this is the question of causal efficacy. But we can also ask whether c's possessing P_i (where P_i is a member of the P_1 class) caused e's being B_i (where B_i is a member of the class of physical tropes, B). This, we are *told*, is the question of causal relevance. But in this case, c's possessing P_i, that event, or event-trope, is identical with c's possessing M_i, where M_i is a member of the mental class M. Thus, it follows that if c's possessing P_i caused e's possessing B_i, then c's possessing M_i caused e's possessing B_i.

But this is *not* the question of causal relevance. The "relevance" question *only* arises if we ask whether that event, or event-trope, c's possessing P_i, that is, c's possessing M_i, succeeded in causing e's possessing B_i *in virtue of* being a P_i-ing or *in virtue of* being an M_i-ing. And this question, it seems, cannot be answered just by appeal to mental/physical tropes, but requires appeal to properties of tropes in virtue of which they are causally efficacious.

The problem can be put more simply like this. Take event c to be the trope P_i, and take event e to be the trope B_i. The question that the trope-physicalist tells us is the question of causal efficacy is the question of what c, that is, P_i, causes. Since P_i just is M_i, if P_i is causally efficacious, M_i is. But there is also a "relevance" question: it is the question whether P_i caused B_i in virtue of being P_i, or in virtue of being M_i. Being P_i and being M_i are properties *of* tropes—where 'property' here means either "trope" or "class of tropes." Minimal physicalism takes these latter properties to be distinct. On either of the two readings of 'property' available, there is a problem of causal relevance. If the reading is "class of tropes," then, for reasons already given, principle MCR is jettisoned. If, on the other hand, the reading is "trope" the problem of causal relevance re-emerges, and the trope strategy does not solve it.

So, neither the "trope" interpretation nor the "class of tropes" interpretation of 'property' works to show how appeal to trope theory solves the problem of the causal relevance of mental properties. Of course, it is open to the trope theorist to deny that tropes have properties. If this option is taken, though, the result will be that the distinction between causal efficacy and causal relevance cannot be made out. And if it cannot, MCR cannot be salvaged.

Thus, irrespective of whether the trope strategy construes the issue of the causal relevance of mental properties to be one that concerns property-types (that is, classes of tropes), or one that concerns tropes alone, it cannot give an effective response to the charge that minimal physicalism leads to the causal irrelevance of mental properties. Its attempts to do so fail to engage with the problem of causal relevance of properties as posed at the outset of our discussion.

IV. APPEALING TO THE PROPERTY EXEMPLIFICATION ACCOUNT (PEA) OF EVENTS

How, then, might one go about attempting to rescue mental properties from the threat of causal irrelevance without rejecting the independently plausible three principles, PCR, EXCL, and CLOS? Our view is that one can do so by appealing to a version of the property exemplification account (PEA), *developed in a particular way*. According to this, events, such as my having pain now, not only have properties, such as the property of being a pain event, but are the exemplifyings of properties, such as the property, has pain. That is to say, they are identical with exemplifyings of (*n*-adic) act-or event properties at (or during intervals of) times in objects.[30] The objects in which such exemplifyings occur are the subjects of those events. And the properties, whose exemplifyings in subjects just are events, are properties, not of events, but of their subjects. For example, the event of my having pain now just is the exemplifying in me of a property of me, the property, *has pain*, now. Such properties are sometimes termed constitutive properties of events, and are so termed because they are the properties of subjects whose exemplifyings in those subjects just are events. So when it is said that events "have" constitutive properties, this is not to be understood as the claim that they possess such properties.[31]

---

[30] In the terminology preferred by Kim, whose version of the account we describe and develop further here, events are *exemplifications* of properties by objects at times. See "Events as Property Exemplifications," in Myles Brand and Douglas Walton, eds., *Action Theory* (Boston: Reidel, 1976), pp. 159–77. But Kim himself, and many others who take a Universalist rather than a tropist view of properties, often use the terms 'instance' as alternatives to the term 'exemplification' (and thus claim, for example, that a mental event is an instance of a property at a time in an object). We ourselves prefer 'exemplifyings' to 'exemplifications' (for reasons akin to those given by Lawrence Lombard in *Events: A Metaphysical Study* (New York: Routledge, 1986)), since it makes clear that events are fundamentally changes, whose "constitutive" properties are dynamic rather than static, or its cognate term, 'instancings'; since we think that failure to do so blurs the crucial distinction between a substance and an event (see Cynthia Macdonald, *Mind-Body Identity Theories* (New York: Routledge, 1989), chapter 4). Given the Universalist (as contrasted with a tropist) view of properties, according to which an exemplification/instance of a property just is the thing that has it; we would have to say that *Jones* is the instance of the property, *runs*, since, according to the property exemplification account, as developed by Kim, this is a property of Jones, and so is a constitutive property of the event which is Jones's running. But although Kim wants to say that the subject of that event is Jones, the exemplification of the property *runs* by Jones is an *event*, a running, not the event's subject. We can avoid this problem altogether if we distinguish instances from instancings (that is, exemplifyings), since we can then maintain (1) that an instance of a property is the thing that has it (whether this is an object or an event), (2) that events just are (that is, are identical with) exemplifyings of dynamic properties of objects

Events construed along these lines are sometimes referred to as "structured particulars." They are deemed so because they have not only constitutive properties, but also constitutive objects (or subjects) and constitutive times. That is to say, it is in the nature of any event to be an exemplifying of a property (of its subject) in a subject at a time. Two conditions on events are essential to the account, one an existence condition and one an identity condition. These are formulated for monadic events as follows:[32]

> *Existence Condition*: Event $[x,P,t]$ exists if and only if the object x has the property P at time t.
>
> *Identity Condition*: Event $[x,P,t]$ is identical with event $[y,Q,t']$ if and only if the object x is identical with the object y, the property P is identical with the property Q, and the time t is identical with the time t',

---

in those objects, *and* (3) that an instance of a property of an event just is the event that has that property. Events, like any other entity, have properties by instantiating them, but their constitutive properties are not, according to PEA, properties that they possess. These distinctions are important to our solution to the problem of causal relevance, since only certain ways of developing the PEA will make that solution possible. For more on the distinction between static and dynamic properties, and the differences between Kim's and Lombard's versions of the PEA, see Cynthia Macdonald, *Mind-Body Identity Theories*, and *Varieties of Things: Foundations of Contemporary Metaphysics* (Malden, MA: Blackwell, 2005).

We now prefer to avoid the term 'instances' entirely, since it suggests a trope view of properties, which we reject. But, since many parties to the dispute concerning the problem of mental causation (for example, Philip Pettit (*The Common Mind: An Essay on Psychology, Society, and Politics* (New York: Oxford, 1993)) and Kim himself (*Supervenience and Mind, Mind in a Physical World*)), regularly talk of events as instances of properties—intending the Universalist view of properties as multiply-exemplifiable entities that can be (wholly) present in many places at the same time—we will, for present purposes, speak in these terms too.

[31] An event's constitutive property can no more be viewed as a property *of* it than its constitutive object can be viewed as a property of it. The claim that P is a constitutive property of e entails, not that P is a property of e, but rather, that being an exemplifying of P is a property of e. Thus, for example, the claim that the property, *firing*, is a constitutive property of the event which is Joe's firing a gun at t entails, not that *firing* is property of that event, but rather, that being an exemplifying of the property, *firing*, is a property of that event.

[32] The exposition of the PEA here is based on this work of Kim's (see especially "Causation, Nomic Subsumption, and the Concept of Event," this JOURNAL, LXX,

where x and y, P and Q, and t and t', are variables ranging over objects, properties, and times, respectively.[33] Descriptions of the form '$[x,P,t]$' are known as canonical descriptions of events because they pick such events out in terms of their constitutive objects, properties, and times.

It is worth noting here that, although Kim himself assumes that events have only one, unique, constitutive property, the version of the account we favor, developed by Lawrence Lombard, allows for an event's having more than one constitutive property.[34] This version is predicated on the assumption that events are paradigmatically and fundamentally changes, where these are *not* to be understood as states or persisting conditions. This assumption is founded on the intuition that some properties are such that their possession by an object at a time implies change, whereas others are not. Lombard labels these two sorts of properties "dynamic" and "static" respectively, and argues that only exemplifyings of the dynamic ones imply the existence of events.

---

8 (April 26, 1973): 217–36, and "Events as Property Exemplifications"). According to Kim, although the first condition is indispensable to the theory, the second, as formulated, is not. The theory could proceed, for example, by defining the predicate 'is an event' over ordered n-tuples of objects, properties, and times. In this case, the ordered triple, $<x,P,t>$, would be an event if and only if x has P at t; and the principles of set theory would guarantee the existence of the triple (assuming, of course, that x, P, and t exist). But Kim himself appears to favor the first method over the second, and it is certainly the preferable one from the point of view of the phenomenon of causal interaction between events, where this is assume to entail their positionality.

[33] More precisely, objects that are *minimal subjects* of events, since, as stated in the text, we prefer the version of the account developed by Lombard in *Events: A Metaphysical Study*. According to this, an object, x, is the minimal subject of an event e if it is the minimally involved subject of e, where the notions of an object's involvement and minimal involvement in an event are defined as follows:

> If x is any object, e is any event, and t is a time, then x is involved in e at t if and only if it is the case that if e occurs (or is occurring) at t, then x changes (or is changing) at t, and a change in x at is identical with e at t; and
>
> If x is any object, e is any event, and t is a time, then x is the minimally involved subject of e at t if and only if (a) x is involved in e at t, and (b) x is the smallest object which is such that a change in x at t is identical with e at t (*Events: A Metaphysical Study*, pp. 122–23).

For more on the details of Lombard's version of the PEA, and his reasons for distinguishing subjects from minimal subjects of events, see *Events: A Metaphysical Study*. That mental and physical events might have different subjects—mental ones having, say, persons, and physical ones having, say, brains—does not preclude identity between mental and physical events on the PEA, since the distinction between subjects and minimal subjects allows for the possibility that persons are not minimal subjects of mental events. For more on this, see Cynthia Macdonald (*Mind-Body Identity Theories*).

[34] See Lombard, *Events: A Metaphysical Study* (especially p. 55.) Lombard points out that the view that events may have more than one constitutive property is *not*-inconsistent with the existence and identity conditions of events as stated by the PEA (though it is inconsistent with Kim's claims on behalf of that account).

If a material substance has a dynamic property during an interval of time, then it will be true that that substance is changing during that interval from having one static property to having another. This will be true because a dynamic property just *is* the property of first having one, then another, static property. Thus Lombard's version of the PEA, unlike Kim's, not only countenances the possibility that an event may have more than one constitutive property, but actually requires it.

The claim that events have constitutive objects, properties, and times, should not be confused with the claim that they are in some way constituted by or composed of objects, properties, and times.[35] In this respect at least, describing events as "structured particulars" is misleading. It invites us to view events as somehow composed of objects, properties, and times, related to each other in something like the way that a chair or any other complex physical object or, perhaps, biological organism is often viewed as composed of or constituted by its parts, arranged in a certain way. But the relationships that the "components" of events bear to one another are very different from the relations that the components of physical things bear to one another. In the case of an event, one component is exemplified by another, at yet another; and it is clear that, whatever the constituents of a biological organism or an artifact may be, they do not bear this relationship to one another.

For this reason, as well as others, talk of events as "structured particulars" is best avoided altogether, and the claim that the components of events are constitutive of them should probably be understood as the claim that events have essentially the structure they have.[36] Indeed, Kim explicitly commits himself to some version of the latter—specifically, to the claim that events are essentially exemplifications of act or event properties at times in objects, and hence, that, for any event, *e*, being an exemplifying of an act or event property at a time in an object is an essential property of *e*.[37] Notice, though, that this would appear to follow from the two basic tenets of the PEA alone. For, irrespective of how the existence condition is interpreted, the mere existence of the relevant *x*, *P*, and *t* are not enough to guarantee the existence of an event. No entity *could* be an event, according to this account, unless it was the exemplifying of a property at a time in a substance.

---

[35] Ehring ("Tropeless in Seattle: The Cure for Insomnia") and Robb ("The Properties of Mental Causation") are both guilty of talk of the "components" of events in this misleading way when discussing the PEA. It encourages a trope conception of properties.

[36] One reason for avoiding such talk is that it encourages the view that the PEA is a "reductive" account of events, a view that Kim himself is quick to deny. Thus, he says, "The account so far presented is not an 'eliminative' or 'reductive' theory of events; that is, it does not attempt to show that events are in some eliminative sense 'reducible' to substances, properties, and times...the account, however, attempts to tell us something about the metaphysical nature of events by relating them to such other ontological categories as substances, properties, and times" ("Events as Property Exemplifications," p. 162).

[37] Kim, "Events as Property Exemplifications," p. 173. Kim is also inclined, though not without qualification, to endorse the claim that each individual event has at least some of its constitutive components essentially, namely, its constitutive object and

As the above discussion suggests; in addition to constitutive properties, events also have characterizing properties. These are properties that events possess, at least some of which they possess in virtue of having the constitutive properties they have. Thus, for example, the event that is my having pain now has the property of being a having of a pain. The event that is the exemplifying of the property, *runs*, in Jones at time *t*, has as its constitutive property a property of Jones. That event has the property of being a running.

Kim takes the constitutive properties of events as being the ones that figure in laws connecting the events that are the exemplifyings of them.[38] His reason may be that it is these properties whose exemplifyings bring about, or cause, other events. However, in the context of our discussion of causal relevance, this claim is extremely odd. For the constitutive properties are properties of the *subjects* of events, not of events themselves. And, as we have construed it, the problem of the causal relevance of the mental concerns the question of whether mental (= physical) events cause the effects they do in virtue of their mental properties, properties of those *events*, such as that of being a thinking of Vienna. And the answer to this depends on a number of issues, one of which concerns the conception of "property" at work in the PEA, and another of which concerns the status of mental properties (of persons). We discuss these issues in turn.

The PEA construes properties as abstract, multiply-exemplifiable entities that can have, but are not identical with, their exemplifyings. According to it, to say that mental events are identical with physical

---

possibly its constitutive property. Whereas the claim in the text says what, in general, events essentially are, this latter claim says what each individual is essentially. We are more directly concerned with the claim in the text.

[38] See Kim, "Causation, Nomic Subsumption, and the Concept of Event," pp. 226–27. Kim's early view of the relation between the mental and the physical is that identity between mental and physical events entails psychophysical laws, and he rejects the minimal physicalism assumed in this paper for that reason. But the PEA does not commit him to that view, and it is part of the purpose of this paper to correct the view that the PEA entails type identity between mental and physical properties whose exemplifyings are events. For more on the details of the PEA and its compatibility with nonreductive physicalism, see Cynthia Macdonald, *Mind-Body Identity Theories*, especially chapters 4–6. (Note that Kim's recent view on the mind-body relation differs substantially from the early one.)

events is to say that each event which is (= is identical with) an exemplifying of a mental property of a subject in that subject at a time is identical with an exemplifying of a physical property of that subject in that subject at that time. Crucially, this amounts to the claim that there is just *one exemplifying* of two properties, one mental, and one physical, by an object at a time.[39] That this is possible is apparent from determinable/determinate examples, such as that of being colored and being red. The most natural understanding of the relation between these properties is that for an object to instance the latter (being red) just is for it to instance the former (being colored): nothing further is required, once the latter is instanced, for the former to be instanced. Unlike the determinable/determinate property relation, the relation between mental and physical properties is not both metaphysical and conceptual. However, if, as seems likely, nonreductive monism is committed to the view that mental properties supervene on physical ones (in a sense shortly to be specified), the result is that mental properties of persons are not themselves constitutive properties of the events that are (identical with) exemplifyings of them, but rather, supervene on those events' physical, constitutive properties (see note 39). And our view is that although the supervenience relation is a weaker metaphysical relation than the determinable/determinate one, both are cases where there can be a single exemplifying of distinct properties.

Thus, appealing to the PEA in order to rescue causal *efficacy* for mental events requires simply recognizing that an event can be a single exemplifying of both a mental property and a physical property. In the case of mental and physical properties of events, we claim that this is just what happens, and hence, that the following *Co-Instantiation Thesis* for events is true:

---

[39] Kim claims on behalf of the PEA that both mental properties (of persons) and physical properties (of persons) are constitutive properties of events, and, in his early work ("Phenomenal Properties, Psychophysical Laws, and the Identity Theory," *Monist*, LVI (1972): 177–92), concludes that token identity theories of the mind-body relation are false, on the grounds that mental properties are not identical with physical ones, but the PEA is *not* committed to this conclusion, for two reasons. First, it requires that an event cannot have more than one constitutive property, and the PEA need not be committed to this (compare our discussion of Lombard's version of the PEA, which rejects it, and note 34). Second, even if one does suppose it, one might claim—as we do—that mental properties of persons supervene on the constitutive properties of physical events, and so are not constitutive properties of those events (see subsequent pages of the text). We prefer this way of reconciling the PEA with token event identity because we think that a proper physicalism must be committed not just to an ontology of physical events, but must also provide an explanation of the relation between mental and physical *properties* which shows them to be, if not physical, not worryingly nonphysical. The first way of reconciling the PEA with token event identity leaves the question of the relation between mental and physical properties completely open.

> (CI) Two or more properties of an event can be co-instantiated in a single instance, that is, there can be just one instance of distinct properties.[40]

By the extensionality of the causal relation, if the physical event is causally efficacious the mental one is.

This shows that the PEA has the resources with which to rescue the causal efficacy of instances of mental properties, which on any plausible account is necessary for the causal relevance of the properties themselves. How does the PEA account for the causal relevance of mental properties? We have seen that, according to it, events have (characterizing) properties, and so exemplify properties as well as being the exemplifyings of properties. Given this, and given the Universalist understanding of properties to which PEA subscribes, whereby an exemplification of a property just is (that is, is identical with) the thing that has (exemplifies) it, exemplifications of mental properties of mental events are identical with exemplifications of physical properties of physical events (since each mental event is identical with a physical event). So, to say that a mental property of a physical event is causally relevant (that is, that a mental event is causally efficacious *qua* mental) is to say *at least* that an exemplification of that property, that is, that event, is causally efficacious in bringing about an effect of that event. This will require that (mental) instance to be a physical instance, that is, will require one and the same event to be an instance of both a mental and a physical property.

We think that this requirement would need to be met anyway, since, on the Universalist conception of properties presumed by the PEA, things exemplify properties, and a thing just is (that is, is identical with) an instance of each property that it has.[41] So, an event exemplifies *all* of its properties, and it is (that is, is identical with) an instance of *every* property that it has.[42] Thus, our solution to the problem of causal efficacy makes *every* property exemplified by an event one whose instancing is causally efficacious in bringing about the associated effects. This result is, in essence, what previous critics of this solution have latched onto: too many properties become causally efficacious, or so it is said.

---

[40] See Macdonald and Macdonald, "Mental Causes and Explanation of Action," and "How to Be Psychologically Relevant," in which this thesis is defended in more detail by appeal to an analogy with biological properties.

[41] Alex Oliver (in "The Metaphysics of Properties," *Mind*, cv (1996): 1–80) speaks as though the Platonic view is the view that universals have instances that are in particulars and bear some kind of relation to their instances. But this cannot be right, since on the Universalist understanding of universals, an instance of a universal just is the thing that has it, that is, a particular concrete object. Thus, for example, an instance of the property, red, just is the red bird. Supposing otherwise encourages the view that instances of universals are tropes, but this is not the Universalist understanding of 'instance'. On the Platonic view, there are no "property instances" that mediate between particular and form, or universal.

[42] In her account of "aspect causation," one similar to ours in its use of property-instances, Laurie Paul suggests that she can develop it either as a trope-view or as in the

We claim that this is an inevitable consequence of the extensionality of the causal relation, and that what critics have conflated are requirements on causal relevance with those on causal efficacy. However, this reply will be unsatisfying unless we can provide an account of causal relevance that makes only some of the properties exemplified relevant to some effects. Since, on our view, mental properties are higher-level ones, the causal relevance of such properties cannot be purchased independently of relations that they bear to certain (that is, physical) lower-level properties. CI, being blind to what more is required beyond causal efficacy for mental properties to be causally relevant, cannot help us to see how such properties can be brought within the domain of causally relevant properties along with lower-level, physical ones, in a way that does not entail that every property of an event is causally relevant to any of its effects. Fortunately, a further thesis is available—one that has already been implicit in our discussion of causal efficacy of instances of mental properties—that applies specifically to higher-level and lower-level properties. Things—objects, events, and other particulars—exemplify properties, but some properties they exemplify *just* by exemplifying others. Consider a red, square box. It has the properties of being red and being square. It also has the property of being colored. It is (identical with) an instance of each property that it has. The box exemplifies the properties of being red, being square, and being colored. It does not exemplify the property of being square just by exemplifying the property of being red. But it does exemplify the property of being colored just by exemplifying the property of being red.[43] Similarly, in the case of mental and physical properties of events, a mental event can exemplify the property, being a thinking of Vienna, just by exemplifying the property, say, being neuro-chemical event α. This view is reasonable independently of the issues surrounding mental cau-

---

PEA. What our development of the PEA suggests is that she will find it impossible to use this approach *and* to isolate the causal relata (property-instances) in the way she does, so as to protect the transitivity of the causal relation. See Paul, "Aspect Causation," this JOURNAL, XCVII, 4 (April 2000): 223–34.

[43] But note that we do not think that the relation between mental and physical properties is a determinable/determinate relation, as standard cases of this relation involve conceptual entailment of the determinable from the determinate property.

sation. Where one property (or properties) of an event is said to realize another property (or properties) of that event, this is by far the most plausible way to construe the relation between the properties exemplified.

Call this thesis the *Property-Dependence Thesis*, to distinguish it from the weaker Co-instantiation Thesis. It states:

> (PD) When properties of events are related as higher-level to lower-level and the lower-level properties realize the higher-level ones, an event exemplifies the higher-level one just by exemplifying the lower-level one.

CI is weaker than PD, since, on the Universalist conception, *all* properties of an event are co-instantiated in a single instance. However, the stronger thesis entails the weaker one: where P is the physical property realizing mental property M, there will be just one instance of both P and M, P_i, that is, M_i.[44]

We are now nearly in a position to see how this understanding of the metaphysics of events can facilitate a reconciliation between MCR and PCR, EXCL, and CLOS, consistently with the view that mental properties are distinct from physical ones. A crucial further claim, noted above, is needed: mental properties whose exemplifyings are mental events are not constitutive properties of those events, but rather, supervene on physical properties constitutive of such events, and, consequently, mental properties of events supervene on physical properties of events. Given the identity condition on events imposed by the PEA, nonreductive monism requires rejection of the view that mental properties are constitutive properties of the events that have them. But, independently of this, the position is committed to some kind of supervenience thesis, since without such commitment it is difficult to fend off the charge that the position is irredeemably dualist because it acknowledges the presence in the natural world, if not of nonphysical events, of free-floating nonphysical properties. Many will think that such a position does not deserve the name "physicalist."

What kind of supervenience thesis best captures the relation between mental and physical and physical properties is a thorny issue, as is well known.[45] Still, for present purposes we can say this much. Take supervenience between the mental and the physical to be that relation which holds between a mental property or set of properties, *M* and another, physical one, *P*, such that any two objects/events indiscernible with respect to *P* cannot diverge with respect to *M*. Further, following Kim, let us distinguish weak from strong supervenience.[46] Then we can define a relation of strong supervenience thus:

---

[44] So, in what follows, when we claim that $P_i = M_i$—that is, that there is one instance of both a mental property and a physical property where mental properties are not identical with, or reducible to, physical ones—we mean more than just that there is just one instance of both a mental property and a physical property (since on the Universalist conception of properties, an event is just one instance of *all* of its properties). We mean that a mental event exemplifies *M just by* exemplifying *P*.

[45] For some sceptical discussion of the value of appeal to psychophysical supervenience, see, for example, Richard B. Miller ("Supervenience Is a Two-Way-Street,"

> (SS) *M*-properties strongly supervene on *P*-properties =*df.* For any possible worlds *w* and *w**, and any individuals *x* and *y*, if *x* in *w* is a *P*-twin of *y* in *w**, and the actual world's laws of physics hold in both, then *x* in *w* is an *M*-twin of *y* in *w**.[47]

where any *x* and *y* are *M* (/*P*) twins if and only if *x* and *y* are exactly alike with respect to their *M* (/*P*) properties.

Understood in this way, our claim is twofold: (1) mental properties of persons supervene on their physical properties, and so (2) mental properties of events supervene on their physical properties. This is consistent with the view that an individual event can be an exemplifying of both a mental and a physical property (of a person), can be an instance of both a mental property and a physical property (of an event), and can be an instance of a mental property just by being an instance of a physical property (of an event).

V. CAUSAL RELEVANCE

What was at stake in the previous section was whether the PEA is consistent with the possibility of mental properties' being causally relevant, where causal efficacy is a necessary but not a sufficient con-

---

this JOURNAL, LXXXVII, 12 (December 1990): 695–701), and Andrew Melnyk (*A Physicalist Manifesto: Thoroughly Modern Materialism* (New York: Cambridge, 2003)). For some examples of work on psychophysical supervenience that seeks to meet objections based on the claim that no satisfactory thesis can be found, see Horgan, "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," *Mind*, CII (1993): 554–86, and Thomas Grimes, "Supervenience, Determination, and Dependency," *Philosophical Studies*, LXII (1991): 81–92.

[46] See Kim, "Supervenience and Nomological Incommensurables," *American Philosophical Quarterly*, XV (1978): 149–56, and "Concepts of Supervenience," *Philosophy and Phenomenological Research*, XLV (1984): 153–76.

[47] This is an adaptation of the definition of strong supervenience given by Brian McLaughlin in "Varieties of Supervenience," in Elias Savellos and Ümit Yalcin, eds., *Supervenience: New Essays* (New York: Cambridge, 1995), pp. 16–59. By *M*-properties (/*P*-properties) we mean the nonempty set, *M* (/*P*), of properties. We choose this version over Kim's principally because it is weaker than his, though his entails it. Kim's implies that it is necessarily the case that if something has an *M* property, then it has some *P* property. But SS could be true if twins had no *P* property at all. It thus allows for the possibility that there might be purely mental worlds. We think this consequence desirable, given that we take physicalism to be true and contingent, and given the variable realizability of mental properties.

dition on causal relevance, given PCR, EXCL, CLOS, and minimal physicalism. It might be thought that although we have shown that PEA is consistent with PCR and CLOS, we have not shown that it is consistent also with EXCL, since we have not established that mental properties and physical properties of events are not in competition with one another. The question that might seem to have remained unanswered is this: Given that mental properties are wholly distinct from physical ones, how can the mere fact, established by PD, that an event can exemplify a mental property just by exemplifying a physical property take the two properties out of competition with one another?

But this question misunderstands the dialectic of the argument of section IV. The argument was not that PD *alone* explains how mental properties do not compete for causal relevance with physical ones. PD was invoked in order to help explain how it is *possible*, given that mental properties are higher-level properties, for them to be causally relevant on the PEA in a way that does not make every property of an event causally relevant to any of its effects—a possibility that CI does not explain. How, then, is PEA made consistent with EXCL given the other principles? Well, look again at what EXCL says. It says that if a property *P* of an event *c* is causally sufficient for an effect *e*, then no other property *Q*, of *c*, *distinct from and independent of P*, is causally relevant for *e*, where a property's being causally sufficient for an effect means that an instance of that property is causally sufficient for that effect, and where a property's being causally relevant for an effect requires that an instance of that property is causally effective in bringing about that effect. What PD tells us is that an event can exemplify—that is, be an instance of—a mental property just by being an instance of a physical property. Given this and CI, PEA is consistent with the possibility that mental properties are causally relevant—instances of mental properties can be causally effective—because they just *are* instances of properties that are causally sufficient. Given supervenience, the mental property's causal relevance is compatible with its subvening physical property's causal sufficiency, and consistent with EXCL's requirements, since mental properties are *distinct from* but not *independent of* physical ones. ("Distinctness" satisfies the "nonreductive" part of nonreductive monism, and we think that any viable physicalism will require dependence.) We do not see how anything more is, or should be, required to show that mental and physical properties need not compete with one another.

But this may invite the question, not of how the PEA is (metaphysically) *consistent* with the causal relevance of the mental—of how a certain metaphysical view of events does not exclude the possibility of the causal relevance of the mental, but of how mental properties can be plausibly seen to be causally relevant. In approaching this matter we issue a prior warning: our approach is deflationary in two ways. First, we take it that a philosophical solution to the problem of mental causal relevance (as well as causal efficacy) is one that demonstrates that such relevance (or efficacy) is unproblematic. The shape of the debate, especially discernible in the writings of Kim, is one in which certain assumptions are meant to ensure a difficulty for *mental* causation. Showing that the difficulty is surmountable is sufficient to dispel the problem. What we have done so far is to take on board the appropriate assumptions and show that the metaphysics of events provided by the PEA can accommodate the causal efficacy of instances of mental properties without any troublesome overdetermination of effects, consistently with the possibility that such properties are causally relevant. The task we are now taking on is that of supporting the claim that a mental event, *qua* mental, can have causal influence, and so can be causally relevant.

Second, we do not aim to provide a general account of causation—though we do think that the metaphysics of events outlined above is important for any such general account. The aim, again, is to demonstrate that the *special* difficulties that some think attach to the possibility of mental causation are not difficulties unique to mental causation. They arise in the context of trying to provide a perfectly general account of the conditions under which it is correct to judge either that *c* caused *e*, or that *c* is causally relevant to *e*. It may well be that the attempt to provide such a general account is itself mistaken, but if that is so it will not be because of any problem specific to mental causation.[48]

We have said that the causal efficacy of *instances* of properties is a necessary condition for the causal relevance of the properties instanced.[49] The issue of causal relevance is one that focuses on the properties, rather than their instances. To see what more is needed, it will be instructive to return to the original source of these debates, Davidson's anomalous monism. Central to Davidson's argument is the *Principle of the Nomological Character of Causality*, the PNCC. This Principle states that events are cause-effect related if and only if there is a law "covering" the causing episode. Putting it in our preferred way, such a law will mention only some of the properties of the events involved in the particular causal affair, those properties of the cause that are "relevant" to the exemplification of certain properties in the effect.[50] For Davidson, only physical predicates can be employed in such laws, leading to the suggestion that only physical properties could be causally efficacious, and hence causally relevant.

---

[48] Hitchcock, for example, thinks that it is futile to attempt to provide a unique characterization of *the* causal relation. See his "Of Humean Bondage," in *British Journal for the Philosophy of Science*, LIV (2003): 1–25. We are sympathetic to this view as far as causal relevance is concerned. We are also inclined to the related view that reductionist accounts are bound to fail.

[49] This is not uncontroversial; Pettit denies it (see *The Common Mind*, pp. 32–42), but we think for no good reason. See our "Beyond Program Explanation," in Geoffrey Brennan, Robert Goodin, Frank Jackson, and Michael Smith, eds., *Common Minds: Themes from the Philosophy of Philip Pettit* (New York: Oxford, 2007).

But it now becomes apparent that the problem of the causal relevance of properties is not one unique to the mental domain. Suppose that we adhere, for the present, to the PNCC requirement on causal interactions. As mentioned above, this privileges only some of the properties of the related events, so it is plausible that some of the *physical* properties of those events will be "nonnomological," or will be specified in laws not germane to a particular cause-effect pair. The law will be applicable because of only certain of the properties exemplified in the cause and its effect. For this causal transaction, then, some *physical* properties will be not relevant—will not be salient. The effect will be deemed to have occurred *qua* only those properties specified in the appropriate law. Or, more precisely, a property exemplified in the effect will be dependent on, or co-vary with, only some property or properties exemplified in the cause. So the problem of the causal relevance of mental properties dissolves into the more general problem of determining which of the properties in a specific interaction are causally significant and which not, no matter whether those properties are physical, chemical, mental, or anything else. Given the generality of this problem, it would be churlish to insist that it be solved for mental causation before a general solution appears.

We can, however, say a bit more about why we think that mental properties are causally relevant, by looking at some examples of causings that are not covered by strict laws. Many recent accounts of causation attempt to extend the scope of causality to cover probabilistic connections between events, and do so by saying that causes "raise the probability" of their effects. There seem to be counterexamples to this: a golfer intends to hit the golf-ball into the hole, but slices horribly, the ball ricocheting off a conveniently situated tree and ending up in the hole. The sequence of events seems to be one in which the slice *decreased* the probability of the ball's ending up in the hole, but nevertheless the slice seems to have been a cause of that effect. The estimates of probability here, we claim, are estimates of the causal relevance of properties, relating the probabilities of sequential exemplifications of properties.[51] Note that no sensible person would want to say that the above example proves that nature is indeterministic. Under some specification of properties, one involving forces acting on a body, what happens when a body in motion impacts on another body at a certain speed and direction, and so on, the probability of the ball's going into the hole *given that confluence of properties* may well be 1, or close to 1. And this combination of forces and resistances surely is involved when the golfer slices the ball. In general, what is "accidental" is relative to a specification of properties. Given that different properties are exemplified in causally related events, the same effect can be accidental (or probability-decreasing) relative to one property-specification, and nonaccidental (or probability-raising) relative to another.[52] This is exactly what one would expect if what is at stake in these discussions is causal relevance, and not causal efficacy.[53]

---

[50] Davidson talks only of predicates, not properties.

This is further illustrated in a recent discussion of probabilistic causation in which Christopher Hitchcock distinguishes between "two effects," a net effect and a component effect.[54] Taking birth control pills containing estrogen has been said to raise the probability of thrombosis occurring while at the same time decreasing the risk of thrombosis by preventing pregnancy. In order to avoid the apparent contradiction, Hitchcock suggests that taking the pills is positive for thrombosis along one causal route (one effect of imbibing estrogen) and negative along another (a second effect, the prevention of pregnancy), where the notion of a "causal route" to an effect is termed a "component effect." Combining the positive and negative component effects will, in this case (restricted to a specific population), yield a net negative effect: the chance of thrombosis (for the relevant population) is lowered.

Hitchcock's "component effect" is akin to our notion of causal relevance, although it needs to be noted that his distinction between net and component effects does not exactly match our causal efficacy/causal relevance distinction. As we understand his distinction, "component effect" is similar to our "causal relevance" insofar as certain properties are privileged as being components of a causal process or as being relevant to a certain effect. However, for us causally relevant properties do not compete with, or dilute the power of, other relevant properties to produce "net effects." Thus, for us causal relevance is always relative to a property *as exemplified* in the effect, whereas for Hitchcock a "component" cause need not be successful in producing an effect. As he puts it, "I assume that at the type-level, the words *cause* more and *prevent* are not success verbs, but rather describe causal tendencies."[55] Further, for him, when an assassin shoots a victim, the assassin's shot has a component effect on the victim's death, but no net effect in those cases where there is a backup who would have shot the victim in the absence of the assassin's shooting. However, for us, the event which is the assassin's shooting is causally efficacious, the property of being a shooting being a causally relevant property of the event.

---

[51] Hitchcock, for example, takes the "factors," such as smoking, that are involved in probabilistic accounts to be "event-types or a property that can be instantiated by the appropriate sort of individual; factors, unlike particular events, are repeatable and may be instantiated on multiple occasions"—"Causal Generalizations and Good Advice," in Henry E. Kyburg, Jr., and Mariam Thalos, eds., *Probability Is the Very Guide to Life* (Chicago: Open Court, 2003), pp. 205–32, p. 210.

[52] This is hospitable to the account given by Hitchcock, which makes a causal claim equivocal unless the probability of the effect given one cause is contrasted with its probability given alternative causes. See Hitchcock, "The Mishap at Reichenbach Falls," *Philosophical Studies*, LXXVIII (1995): 237–91.

[53] So there is a lesson to be learnt from these debates about the causal relevance of mental properties, and that is that recent discussions of causation *simpliciter* need to pay more attention to these metaphysical foundations. Many of these discussions are careless about the distinction between causal efficacy and causal relevance.

[54] Hitchcock, "A Tale of Two Effects," *The Philosophical Review*, CX (2001): 361–96.

Given that events exemplify diverse properties, it will be true that only some of the properties of the cause will be relevant to some of the properties of the effect. Which properties are relevant in any particular case will be discovered empirically, and, so far as we can see, there is nothing to prevent the discovery that mental properties are causally relevant. Of course, even when exemplified, they may not be relevant to some effect-properties, but they are not unique in this respect. Some properties exemplified in nonmental causings are not relevant to some of their effect-properties either, this just being a consequence of *any* metaphysics of events that allows for an event to be an exemplification of many properties.

Consider, for example, a window that shatters when a rock of 5 lbs or more impacts on it at 10 mph or more, and so shatters because it is struck by a 7 lb rock thrown at 10 mph. It is causally relevant that the rock weighs over 5 lbs, even though this instance of weighing over 5 lbs is also an instance of weighing over 2 lbs. The throwing of a rock weighing over 2 lbs will not explain the window's shattering, though *in this case* the instance of the property, weighing over 2 lbs, just is the instance of weighing over 5 lbs (= the instance of weighing 7 lbs), and so was causally implicated in the shattering. This property, weighing over 2 lbs, does not help explain the shattering because other instances of it in rock-throwings will not be ones in which the window will shatter. But this truth about other instances is irrelevant to the causal efficacy of this instance, it being pertinent only to the matter of causal relevance. So objections to our account that protest that it makes *too many* properties causally relevant simply confuse causal efficacy with causal relevance.

---

[55] Hitchcock, "A Tale of Two Effects," p. 375, fn. 9. Where Hitchcock talks of "component effects" as "causal tendencies" we would prefer to talk of the causal power of a property, the property's causal relevance being dependent both on having such a power and being instanced (exemplified).

There remains, however, a potentially serious objection, one stemming from the role played by the PNCC. Critics may agree that mental properties will not be unique in sometimes being causally irrelevant, but insist that, given the PNCC *and* the anomalousness of the mental (that is,that mental properties do not figure in causal laws), they are unique in *never* being causally relevant. That they might, on particular occasions, join other, physical, properties in being causally irrelevant will be of little comfort if they are, by their nature, *never* relevant.

But this will be so only if we concede two claims. The first is that the mental is anomalous, and the second is that causal relevance can be secured only by nomologicality. Both claims can be and have been contested. Some have modified the first claim by distinguishing between "strict" and "loose" laws, making much of the use of "*ceteris paribus*" laws in the special sciences. It would be open to a proponent of the PNCC to loosen one's understanding of "law" so as to include these less-strict laws in the "N" part of the PNCC, thus still allowing that Principle the role of sole bestower of causal relevance. As long as mental properties were to figure in these loose laws, that would permit their causal relevance.

There may well be ways of pursuing this strategy so as to allow for mental causal relevance, in the process diminishing the tension between the PCR and MCR. But with this tension in mind, we prefer to challenge the second claim, that the PNCC is the sole guarantor of the causal relevance of properties. First, it is worth noting that the principle does not serve this function in Davidson's original argument; it serves as a guarantor of causal efficacy, not of relevance. More interestingly, one might view the PNCC as providing *a* criterion for one kind of causal relevance, *physical property* causal relevance. This makes the role of the PNCC more extensive, in ensuring the causal efficacy of mental events in a way that secures the causal relevance of the physical properties exemplified by those events.[56]

---

[56] This extended role serves to integrate mental causation more thoroughly with physical causation. For further remarks on this integration (on how harmony of diverse causal relevances can be achieved), see Graham Macdonald, "Reduction and Evolutionary Biology," in *Reduction, Explanation and Realism,* Kathleen Lennon and David Charles, eds. (New York: Oxford, 1992), pp. 69–96. One further feature of such integration is the avoidance of troublesome "downward" causation. On how we avoid that trouble, see "Beyond Program Explanation."

Second, and relatedly, the PNCC does not serve the function of being the *sole* guarantor of causal relevance in contemporary controversies about causation, as the briefest of summaries of some of this literature will show. It has been suggested that many theories of causation have the following form:

> (CF) *C* has an effect on *E* iff *E* depends upon *C*—that is, iff *E* varies as *C* is varied while holding fixed other appropriate factors.[57]

This can be filled out in various ways. One account follows David Lewis and interprets causal dependence in terms of counterfactual dependence.[58] Given the difficulties of making this work satisfactorily,[59] Lewis has recently suggested a requirement of "counterfactual co-variation."[60] This requires, minimally, that variations in the manner of *c*'s occurring co-vary with changes in the manner that *e* occurs. Thus, for example, the atmospheric pressure causes the barometer pointer's position to be what it is because variations in atmospheric pressure co-vary with changes in the position of the pointer.[61]

It is important that the notion of "variation" here only makes sense once properties enter the picture, since what is seen as crucial is the existence of *some* patterned variation between a property (or properties) exemplified in the cause and a property (or properties) exemplified in the effect.[62] The proponent of the possibility of the causal relevance of mental properties will insist, justifiably, that this is satisfied in the paradigmatic cases of mental causation. Thus, suppose that an officer shouts an order, and the soldier obeys it by standing to attention. Had the shooting not occurred, the standing to attention would not have occurred. Had a different order occurred, a different action would have followed.

---

[57] This is Hitchcock's formulation, in "A Tale of Two Effects," p. 369.

[58] See the papers on causation in Lewis, *Philosophical Papers*, Volume II (New York: Oxford, 1986). Lewis's more recent elaboration of his theory is in "Causation as Influence," this JOURNAL, XCVII, 4 (April 2000): 182–97.

[59] For some of the discussion see the papers collected in John Collins, Ned Hall, and Laurie Paul, eds., *Causation and Counterfactuals* (Cambridge: MIT, 2004).

[60] See Lewis, "Causation as Influence." See also Hitchcock, "The Intransitivity of Causation Revealed in Equations and Graphs," this JOURNAL, XCVIII, 6 (June 2001): 273–99.

[61] Another account along similar lines is James Woodward's interventionist account: "as a rough approximation, a necessary and sufficient condition for X to cause Y or to figure in a causal explanation of Y, is that the value of Y would change under some intervention on X in some background circumstances..." (*Making Things Happen* (New York: Oxford, 2003), p. 15). The qualifications concerning *some* intervention in *some* background circumstances signal the distance we have come from strict (or even looseish) laws. Woodward's account is significant in being nonreductionist.

[62] This is apparent in the literature on causal models, where a causal model consists of an ordered pair $<V, E>$ where V is a set of variables and E is a system of equations relating those variables. The values of the variables are best seen as properties of an event, ones that can vary in different ways: "For example, the values of a variable could represent the mass of some object, or they could represent various alterations of a particular event"—Hitchcock, "The Intransitivity of Causation Revealed in Equations and Graphs," p. 280.

Now, it might be objected that the systematic co-variation displayed in the above examples is the result of what may be called the "rational connectedness" of the events (what used to be called "conceptual connections" between the events), rather than simple (strict or loose) nomological dependence.[63] Suppose that it is true that the connections displayed are "rational"; does it matter? It does if one's aim is to argue for the *a priori* irreducibility of the mental to the physical, as one may then wish to stress the difference between the patterns displayed in virtue of rational connections and those displayed in virtue of "mere" causal connections. But however important the difference is for those purposes, it is not important in the context of establishing the possibility of mental properties' being causally relevant. The "monist" part of anomalous monism ensures that the rational connections are also causal connections, so, short of denying this possibility (that reasons are causes), there is no good reason to affirm the impossibility of the causal relevance of mental properties.

It may also be worth pointing out here that recent disputes about causation, particularly those concerning the transitivity of the causal relation, make unapologetic use of examples that involve rational connections. Here is an example, one of many. *Dog Bite* is a little narrative in which we have a Terrorist (*T*) wanting to explode a bomb, needing to push a detonator-button (*P*) at noon in order for the bomb to explode.[64] Being right-handed, *T* would have pushed the button with his right forefinger (P_r), but a dog bites that finger off (*D*), so event *P* is a left-forefinger pushing (Pl), which results in the explosion (*E*) occurring as planned. This seems to be a situation in which *D* causes Pl, and P_l causes *E*. Does *D* cause *E*, as transitivity would have it? It appears counterintuitive to say that the dog bite caused the explosion. The reason is that, given *T*'s wishes, and given the means at *T*'s disposal (left forefinger, feet, elbows, and so on), it is just plausible (reasonable) to assume that *T* will be intelligent enough, and determined enough, to carry the plan through regardless of the dog bite.[65] The "rational cause" of the explosion is undisturbed by the dog bite.

---

[63] In "How to Be Psychologically Relevant," we defended the view that causal relevance is dependent on *patterns* in nature, rational connections displaying one such pattern.

[64] The example is raised in Michael Mcdermott, "Redundant Causation," *British Journal for Philosophy of Science*, XLVI (1995): 523–44.

[65] An argument for transitivity is given by Hall, "Causation and the Price of Transitivity," this JOURNAL, XCVII, 4 (April 2000): 198–222. An argument against is given by

Whether these cases support intransitivity or not is not our concern. What is to the point is that these examples are ones where the issue is one of causal relevance, of whether the property (-ies) instantiated in the cause are causally relevant to the property (-ies) instantiated in the effect. And the kind of systematic dependence between (exemplifications of) these properties that legitimizes causal relevance is not restricted to nonrational dependencies, as *Dog Bite* shows.

So far, so good: the examples above establish that putative cases of mental causation satisfy the test for causal relevance thought appropriate by contemporary theorists, that of systematic property dependence or co-variation. But the lingering doubt will be that we have not done what we promised, which is to show that there need not be any tension between physical and mental causal relevance, between the MCR and the PCR. It should be clear, however, how the supposed tension dissipates. Even if nomologicality were the only guarantor of causal relevance it would remain possible that two causally related events would fall under two (or more) laws, the property (properties) cited in the antecedent of law 1 being relevant to the exemplification of the property (properties) cited in its consequent, and the same for law 2. In this case we would have two relevance relations, and there is simply no reason to see them as competing, so no reason to say that one must preclude the other. Overdetermination has to do with causal instances—efficacy, not relevance.

And again, there is no reason to see these different co-variation relationships as excluding one another. At least they do not compete *simply in virtue of being different properties.* They can co-exist harmoniously. In the mental case, this is guaranteed by the *Property Dependence* thesis (PD) and the other principles at play that together ensure that the physical properties of an event that realize a mental property will not conflict with the causal power of the mental property.

Any remaining doubt may take this dependency to be the problem rather than part of the solution. It may be thought that PD ensures that the physical property/ies will eclipse the relevance of the dependent mental property. But this is not true: it is because PD is weaker than reduction—variable realization is accepted—that mental causal relevance remains. The systematic co-variation between exem-

---

Hitchcock, "The Intransitivity of Causation Revealed in Equations and Graphs." At one point Hitchcock goes explicitly rational: "Given that he did not push the button with his right hand, why should the dog bite make any difference to whether he pushes it with the left? He wanted the bomb to explode, would he not push the button with his left hand regardless of whether the dog bit his right?" p. 292.

plified mental properties is not a simple "upward reflection" of what happens at the level of the realizing physical properties. For it to be such an "upward reflection" would require the identity of mental and physical properties, and it would need an independent argument to support any such identity claim. We maintain that the identity claim cannot be supported by arguments using the causal irrelevance of the mental as an essential premise; not without begging the question.[66]

Causal relevance, then, is ubiquitous, and would remain so even if the world were shorn of higher-level properties. The causal relevance of a property, *P*, requires at least that instances of *P* are causally efficacious and that there are objective patterns, possible systematic co-variations among properties, in which *P* participates. Mental properties can satisfy these conditions, and that the pattern exemplified between, say, an intention and an action is one of rationality, or intelligibility, does not preclude the causal relevance of the intention to the performance of that action. So it turns out that the causal relevance of mental properties is no more mysterious than is causal relevance in general.

VI. CONCLUSION

The problem of mental causation divides into the problem of the causal efficacy of mental events and the problem of the causal relevance of mental properties. We have argued that the former problem can be satisfactorily settled by adopting either a metaphysics of tropes or a version of the Property Exemplification Account (PEA) of events, but only the PEA can provide the foundations for an account of mental causal relevance. Regarding the latter we have claimed that a property, *P*, is never causally relevant without there being another property relative to which it is causally relevant, so the question of which properties of events are the causally relevant ones is not one that can be answered *tout court* and independently of the types of effects that they bring about. This is a general difficulty of determining which properties of a cause are relevant to which properties of an effect, not one specific to mental causal relevance.

This suggests that concerns arising from *Exclusion* (EXCL) are overstated. For EXCL states that if there is a property, *P*, whose instance in an event *c* is causally sufficient for an effect *e*, then there can be no other distinct and independent property, *Q*, that is causally relevant for *e*. And CLOS states that if any physical event has a cause, it has a sufficient physical cause, whose physical properties are causally sufficient for its effect. However, appeal to the physical properties of physical events does not, given this, thereby rob mental properties of physical events of either their causal efficacy or their causal relevance, and so their causal-explanatory power remains intact. Causal efficacy is retained because the *Co-instantiation Thesis* (CI) permits the co-instantiation of the physical and mental properties, and supervenience and the *Property-Dependence Thesis* (PD) provide compelling reasons for assuming that they are so co-instantiated.

---

[66] Kim employs an argument using a "causal inheritance principle"; we address that argument in "Beyond Program Explanation." Some of what Karen Bennett argues for in "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract it" (*Noûs*, XXXVIII (2003): 471–97) is relevant here, especially sections 7–10, but our approach differs crucially from hers in that we accept psychophysical event-identity whereas she rejects it.

PD also provides a way of defusing the threat to causal relevance arising from EXCL, since it provides a way of seeing how it is that the mental properties are not independent of the physical properties of such events. It has been thought that such a dependence would lead to causal irrelevance, or that the dependent higher-order properties are not doing the "real" causal work.[67] What our solution does is show how mental properties can be irreducible (and hence independent in one way) whilst still being causally efficacious and causally relevant. The causal relevance of mental properties requires the causal efficacy of their instances plus participation in a pattern (a system of potentially co-varying properties), and participation in the rationalistic pattern, along with the Co-instantiation Thesis (CI) and PD, purchases this consistently with CLOS. So PD, combined with the PEA, seems to be a viable way of resolving the putative conflict between MCR, PCR, EXCL, and CLOS, given minimal physicalism.

CYNTHIA MACDONALD
GRAHAM MACDONALD

Queen's University, Belfast
University of Canterbury, Christchurch

---

[67] David Braddon-Mitchell provides an expression of this common view: "...supervenient causation is best taxonomised as explanation, and causation is best reserved for whatever basal transactions actually perform the work in the universe"—"The Microstructural Causation Hypothesis," *Erkenntnis*, XXXIX, 2 (1993): 257–83, on p. 279. The popular "Program Model" of how higher-order causes relate to their lower-order counterparts suggests the same position: "The general idea in the program model...is that a higher-order property is causally relevant to something when its instantiation ensures or at least probabilifies, in a non-causal way, that there are lower-order properties present which produce it"—Pettit, *The Common Mind*, p. 37.

COMMENTS AND CRITICISM

COMBINATION, CONVENTION, AND POSSIBILITY*

Roberta Ballarin suggests that Saul Kripke in *Naming and Necessity* adopts two incompatible modal conceptual schemes together.[1] Ballarin's bold and serious charge is mistaken, I think. Nevertheless, it offers a welcome opportunity to clarify Kripke's position and the range of available positions on some basic questions about relations of possibility to combination and to convention.

Ballarin argues that *Naming and Necessity* contains "not one, but two conceptual frameworks" and that, since "there is an inherent tension between the two schemes," "we have to make a choice" between them (620). By "inherent tension" she means logical inconsistency. The views Kripke adopts with one scheme, she says, "contradict" the views he adopts with the other scheme (622).

I will selectively consider only what Ballarin's complex schemes say about the extent to which consistency of object-property combinations is ever sufficient for possibility and the extent to which mere stipulation ever determines the truth values of modal property attributions. I will describe a range of conceptual schemes other than Ballarin's that engage those questions.[2] I will argue that one par-

---

* Versions of the present paper were read at the 2003 Northwest Conference on Philosophy at Reed College and at the 2005 Spring Philosophy Colloquium at Rhodes University. I thank Roberta Ballarin, Luke Buckland, David Chalmers, Dean Chapman, Louis deRosset, Mark Leon, Thaddeus Metz, Guy Rohrbaugh, and Ori Simchen for helpful comments and discussions.

[1] Ballarin, "The Interpretation of Necessity and the Necessity of Interpretation," this JOURNAL, CI, 12 (December 2004): 609–38; Kripke, *Naming and Necessity*, revised and enlarged edition (Cambridge: Harvard, 1980), hereafter NN.

[2] I suppose the following. A *conceptual scheme* on a given set of theses is an assignment, perhaps incomplete, of attitudes (affirmation, denial, or suspension of judgment) to the individual theses in the set. (I treat affirmation, denial, and suspension of judgment as distinct attitudes in order not to foreclose liberal views about what combinations of attitudes are psychologically possible.) To *adopt* a scheme is to have an attitude towards a thesis if the scheme assigns the attitude to the thesis; and, other things being equal, adopting one scheme on a set of theses is prima facie *preferable* to adopting a second scheme on that set if attitudes are assigned to the theses in the set more completely, or more rationally, or with more justification by the first scheme than by the second scheme. And a set of theses *engages* a set of questions iff the set of theses logically partitions the space of answers for each question in the other set, while a scheme's engagement of a set of questions is derivative from that of the set of theses on which it is a scheme and varies with its completeness as a scheme on that set of theses.

0022-362X/06/0311/577–86

ticular scheme is at least provisionally preferable to the others and Kripke in *Naming and Necessity* adopts that scheme.

I

Each of Ballarin's two schemes affirms its own thesis about the relation of possibility to combination, which Ballarin expresses as follows.

> *Thesis* (*a1*): "All...consistent [object-property] combinations are possible" (626).
>
> *Thesis* (*a2*): "[C]onsistency [of object-property combinations] is not in itself sufficient for possibility" (636).

Ballarin does not define "consistent object-property combination" but I suppose the following. Given a set of objects and a set of properties, an *object-property combination* is an assignment of properties to objects that says, for each object and each property in the given sets, whether the object has or lacks the property; a *consistent* object-property combination is one that does not contradict itself by saying, for some object and some property, both that the object has the property and that the object lacks the property; and an object-property combination is *possible* iff there is a possible world in which, for every object and every property in the given sets, the object has the property if the assignment says the object has the property and the object lacks the property if the assignment says the object lacks the property. With this understanding Ballarin aptly calls (a1) a form of "combinatorialism" and I will also call (a2) a form of essentialism (622).

Now, Ballarin thinks that we have to choose between (a1) and (a2) but that Kripke affirms both. It does appear that (a1) and (a2) cannot both be true. But why does Ballarin think that Kripke affirms both? The answer is that Ballarin ostensibly finds the combinatorialism of (a1) in the dice passage and the essentialism of (a2) in the passages about necessity of origin and substantial make-up in *Naming and Necessity* (620–26, 631–34). In the dice passage Kripke says that all consistent combinations of two ordinary dice (members of a set of objects, call it *D*) and the six possible results of the throw of a single die (members of a set of properties, call it *F*) are possible, that is, each consistent object-property combination (from *D* and *F*) is realized in one or another of thirty-six "miniworlds" (NN 16). But Kripke intends the "analogy" of the dice to "clarify" his view of possible worlds generally (NN 16). So, Ballarin reasons, Kripke affirms that (a1) all consistent object-property combinations whatever are possible, that is, each combination is realized in some possible world or other. In the passages about necessity of origin and substantial make-up, on the other hand, Kripke says, for example, that this wooden table could not have been made of ice (NN 113–14, 141–42). The combination of this table with the property of having been made of ice, though consistent, is not possible. So, Ballarin reasons, Kripke affirms that (a2) consistency of object-property combinations is not sufficient for possibility.

But Ballarin is wrong about the dice passage. Kripke does not affirm (a1) there, but only a weaker thesis that is consistent with (a2).

To show that, it will help first to express (a1) and (a2) as explicitly as possible. While the notion of an object-property combination is defined relative to given sets of objects and properties, Ballarin's expressions of (a1) and (a2) do not give sets of objects and properties from which combinations are drawn. Clearly, though, Ballarin intends combinations to be drawn from among all objects and all properties. Her meaning is both clear and explicit if (a1) and (a2) are respectively expressed as follows.[3]

> *Strong Combinatorialism*: For every set of objects and every set of properties, every consistent object-property combination from those sets is possible.
>
> *Weak Essentialism*: For some set of objects and some set of properties, some consistent object-property combination from those sets is not possible.

These explicit expressions of (a1) and (a2) help to clarify a number of aspects of Ballarin's and Kripke's views. First, Strong Combinatorialism and Weak Essentialism are logical contradictories. So, Ballarin is right to think that we have to choose between (a1) and (a2). Second, Kripke affirms Weak Essentialism, offering the table example and others as "examples of essential properties" (NN 115). So, Ballarin is right to think that Kripke affirms (a2). Third, as Kripke says, Weak Essentialism (a2) is intuitively correct: "[S]ome things...we don't think could have been otherwise" (NN 41). Fourth, since (a1) and (a2) are logical contradictories and (a2) is intuitively correct, Strong Combinatorialism (a1) is intuitively incorrect. Finally, Kripke denies the following thesis, which is not entailed by Weak Essentialism.

---

[3] Where U is the set of all objects and Φ is the set of all properties, Ballarin herself might offer the following somewhat different explicit expressions of (a1) and (a2) (see 613–14, especially note 14, and 622).

> *Maximal Combinatorialism*: Every consistent object-property combination from U and Φ is possible.
>
> *Minimal Essentialism*: Some consistent object-property combination from U and Φ is not possible.

If U and Φ do both exist, then Maximal Combinatorialism and Minimal Essentialism are respectively equivalent to Strong Combinatorialism and Weak Essentialism. If U and Φ do not both exist, then Maximal Combinatorialism is true and Minimal Essentialism is false, each vacuously and uninterestingly, but the more interesting theses of Strong Combinatorialism and Weak Essentialism are not equivalent to them. Now, it is controversial whether U and Φ exist. (If sets are objects, for example, but there is no set of all sets, then U does not exist.) But those controversies are not relevant here, so I will identify (a1) with Strong Combinatorialism and (a2) with Weak Essentialism.

> *Strong Essentialism*: For every set of objects and every set of properties, some consistent object-property combination from those sets is not possible.

Strong Essentialism says in other words that every object essentially has or lacks every property it has or lacks.[4] Strong Essentialism is intuitively incorrect, as Kripke says: "[W]e do think some things could have been otherwise" (NN 41).

Does Kripke affirm (a1), though, as Ballarin says? No. Kripke affirms the following thesis, which does not entail Strong Combinatorialism (a1) and which is consistent with Weak Essentialism (a2).

> *Weak Combinatorialism*: For some set of objects and some set of properties, every consistent object-property combination from those sets is possible.

As noted, Kripke denies Strong Essentialism, the negation of Weak Combinatorialism. And he affirms Weak Combinatorialism in the dice passage, where he is concerned to draw attention to certain intuitions about the real possibilities presupposed by "school exercises in probability" (NN 16). The example of the dice and their face-showing properties shows, among other things, the intuitive correctness of the following thesis.[5]

> *Ordinary Dice Combinatorialism*: Every consistent object-property combination from D and F is possible.

Weak Combinatorialism follows immediately from Ordinary Dice Combinatorialism. But the dice example provides no support for

---

[4] Strong Essentialism is equivalent to *Robust Essentialism*, the thesis that every object essentially has or lacks every property it has or lacks. Proof: (Left to right.) Let a be an object and G be a property. There are exactly two consistent object-property combinations from $\{a\}$ and $\{G\}$: C1 says a has G; C2 says a lacks G. Suppose a has G. Then C1 is possible and, by Strong Essentialism, C2 is not possible. So, a essentially has G. Similarly, if a lacks G, then a essentially lacks G. (Right to left.) Let O be a set of objects and P be a set of properties where a is an object in O, G is a property in P, C1 is a consistent object-property combination from O and P that says a has G, and C2 is a consistent object-property combination from O and P that says a lacks G. Suppose a has G. Then, by Robust Essentialism, a essentially has G. So, C2 is not possible. Similarly, if a lacks G, then C1 is not possible. QED

[5] Strictly speaking, it is only with qualifications that precisely thirty-six object-property combinations from D and F are consistent and possible. Kripke mentions that "we (fictively)...ignore the fact that one or both dice might not have existed" (NN 16). But we also ignore, for example, the fact that one or both dice might be deformed by the throw so as to show more or fewer than one number face-up each.

Strong Combinatorialism, since it provides no support for the following thesis, which is entailed by Strong Combinatorialism.

> *Strong Dice Combinatorialism*: For every set *I* of properties, every consistent object-property combination from *D* and *I* is possible.

Strong Dice Combinatorialism is not entailed by Ordinary Dice Combinatorialism. So, neither Strong Dice Combinatorialism nor Strong Combinatorialism is supported deductively by the dice example. Moreover, Strong Dice Combinatorialism is intuitively incorrect. Intuitively there is no miniworld in which die A has the property of being a musical tone, or a prime number, or a shadow, or a boundary. Intuitively not every set of objects and every set of properties are relevantly similar to *D* and *F*. So, neither Strong Dice Combinatorialism nor Strong Combinatorialism is supported nondeductively by the dice example.

Nor can Kripke reasonably be held to think that the dice example does support Strong Combinatorialism. He says that, unlike possible worlds generally, "the miniworlds...are tightly controlled, both as to the objects involved (two dice), the relevant properties (number on face shown), and (thus) the relevant idea of possibility" (NN 18). So, he clearly does not intend the analogy between miniworlds and possible worlds to support the conclusion that possible worlds generally have every characteristic of miniworlds, including combinatorial exhaustiveness of some strong sort. On the contrary, he intends the analogy only to "defend...the *propriety* of giving possible worlds in terms of certain particulars as well as qualitatively, whether or not there are in fact qualitatively identical but distinct worlds" (NN 18 note 17). Intuitively there is no impropriety in giving "the [miniworld] (die A, 6; die B, 5)...as such," even if that miniworld is qualitatively indistinguishable from the distinct miniworld (die A, 5; die B, 6) (NN 17). Analogously, Kripke thinks, there is no impropriety in giving the possible world, or perhaps rather the set of possible worlds (Aristotle, fond of dogs) as such, whether or not Aristotle is qualitatively indistinguishable from some other person in some other respect (NN 6–15). Kripke's analogy thus is served by Weak Combinatorialism, which provides qualitatively indistinguishable but numerically distinct and referentially discriminable miniworlds. But the analogy does not require or support Strong Combinatorialism.

II

Besides affirming (a1) and (a2) respectively, each of Ballarin's two schemes affirms its own thesis about the relation of possibility to convention, which I will express as follows.

> *Strong Conventionalism*: Every modal property attribution that has a truth value is made true or made false merely by stipulation. (Ballarin calls this thesis "Stipulability of Modal Predicates" (621).)
>
> *Strong Realism*: No modal property attribution that has a truth value is made true or made false merely by stipulation.

Granted, we do not have to choose between Strong Conventionalism and Strong Realism, since they are not logical contradictories but only logical contraries. They might both be false and, if they are both false, then both of the following theses are true.

> *Weak Conventionalism*: Some modal property attribution that has a truth value is made true or made false merely by stipulation.
>
> *Weak Realism*: Some modal property attribution that has a truth value is not made true or made false merely by stipulation.

But our choice is constrained nevertheless, since Strong Conventionalism and Strong Realism cannot both be true. We have to choose among the alternatives of affirming only one, or only the other, or neither.

Now, Ballarin thinks that Kripke does affirm both Strong Conventionalism and Strong Realism. Ballarin ostensibly finds Strong Conventionalism in the dice passage and Strong Realism in the passages about necessity of origin and substantial make-up. She imputes Strong Conventionalism to Kripke on the basis of his talk in the dice passage of miniworlds being "*given* as such" (621, NN 17). Though the two theses themselves are independent, she says that his affirmation of Strong Combinatorialism rests on his affirmation of Strong Conventionalism. "In calculating 36 possible outcomes for [dice] A and B, Kripke relies on a stipulative maximum principle for possibility—any combinatorial arrangement of the dice is deemed possible" (622). Generalizing from miniworlds to possible worlds, Kripke's one view is the following according to Ballarin.

> We take some given objects and properties.... Next, we apply the Stipulability of Modal Predicates principle—all the combinations of object-property pairs are declared possible. Notice that in proclaiming all such states as possible, we proceed purely a priori, before we know anything more—actual empirical facts—about the objects and the properties in question (622).

Kripke's well-known dictum that "'Possible worlds' are *stipulated*" thus means, on Ballarin's account, that mere stipulation makes possible worlds to be (NN 44). At the same time, Ballarin imputes Strong Realism to Kripke on the basis of what she sees as a common pattern or mechanism in the examples in the passages about necessity of origin and substantial make-up.

> [The] form of argument...is an instance of *modus ponens*:
>
> $$\begin{array}{l} P \\ P \rightarrow \Box P \\ \hline \Box P \end{array}$$
>
> We may call this mechanism "Kripke's Mechanism of Necessity Projection" (in short, KM). We are given a nonmodal truth P about a given subject, for example, that this table is made of wood. Given the particular identity of the individuals and properties involved in P, we have ground to necessitate P. We then detach and conclude: P is necessary (631).

Generalizing from the examples, Kripke's other view according to Ballarin is that, in every case of a modal truth, it is not mere stipulation but "The actual nonmodal truth [about the individuals and properties involved that] supplies the primal basis [and modal truths] are determined from this primal basis" (630).

But Ballarin is wrong again about the dice passage. It counts against her imputation of Strong Conventionalism to Kripke that he nowhere explicitly affirms either Strong or even Weak Conventionalism and nowhere rebuts W.V. Quine's controversial but well-known and influential argument against conventionalism about necessity.[6] Moreover, when Kripke speaks in the dice passage of miniworlds being "*given*," he is reasonably understood as relying only on what Ballarin calls "the innocent principle of The Stipulability of the Subject," which "concerns the subjects of the discourse" (621, 622). Kripke only stipulates "what individuals [and properties] he is talking about, namely, [dice] A and B [and their face-showing properties]" (621). But he does not "[make] use of the exceedingly strong principle of the Stipulability of Modal Predicates," according to which mere stipulation makes-true the possibility of a die's showing a particular face after a throw (622). On the contrary, he says that "The thirty-six possibilities...are...*states* of the dice" (NN 17). Kripke can reasonably be understood to think of the dice and the possibilities of their combinations in just the way he thinks of this table and the necessities of its origin and substantial make-up. In particular, the possibilities of the dice's combinations are known a posteriori, by what I will call

---

[6] Quine, "Truth by Convention," in his *The Ways of Paradox and Other Essays*, revised and enlarged edition (Cambridge: Harvard, 1976), pp. 77–106.

Kripke's *Mechanism of Possibility Projection* (KMP). KMP is the following instance of modus ponens.

$$\frac{\begin{array}{l} P \\ P \rightarrow \Diamond Q \end{array}}{\Diamond Q}$$

For example, let *P* abbreviate "these dice are fair (regular in density, shape, and texture, and so on)" and let *Q* abbreviate "these dice will land A-*m* and B-*n* after a throw" (where *m* and *n* are numbers between 1 and 6). The possibility of these dice's landing A-*m* and B-*n* after a throw thus is grounded in the nonmodal truth about the dice and not in mere stipulation, just as the necessity of this table's being made of wood is grounded in the nonmodal truth about the table and not in mere stipulation. Generalizing from miniworlds to possible worlds, Kripke is reasonably understood as thinking that "'Possible worlds' are *stipulated*" only in the innocent sense that we can stipulate the subject of our discourse when we are talking about counterfactual situations.[7]

Ballarin is also wrong this time about the passages about necessity of origin and substantial make-up. Those passages' examples, and the dice example properly understood, establish the intuitive correctness of Weak Realism and the intuitive incorrectness of its negation, Strong Conventionalism. However, no mere accumulation of individual examples entails Strong Realism and our intuitions do not go so far, either on their own or with the aid of any uncontroversial theory, as to rule out Strong Realism's negation, Weak Conventionalism. So, suspension of judgment about the two is worth considering. Now, while Kripke nowhere explicitly affirms Weak Conventionalism, he nowhere explicitly denies it either. Imputation to Kripke of either affirmation or denial of Weak Conventionalism thus would be at odds with his own silence and with the unsettled character of our relevant intuitions and theory. So, Kripke is reasonably understood as suspending judgment about Weak Conventionalism and Strong Realism.

III

The Strong and Weak versions of Combinatorialism, Essentialism, Conventionalism, and Realism jointly engage some basic questions about relations of possibility to combination and to convention: To what extent is consistency of object-property combinations ever sufficient for possibility? To what extent does mere stipulation ever determine the truth values of modal property attributions?

---

[7] Kripke says as much of his dictum. See also Nathan Salmon, "Trans-World Identification and Stipulation," *Philosophical Studies*, LXXXIV, 2–3 (December 1996): 203–23.

Ballarin's two conceptual schemes on the relevant theses are unappealing. Ballarin's first scheme affirms both Strong Combinatorialism and Strong Conventionalism. But Strong Combinatorialism and Strong Conventionalism are intuitively incorrect, so denial of each is prima facie justified. Ballarin's second scheme affirms both Weak Essentialism and Strong Realism. Granted, Weak Essentialism is intuitively correct. But Strong Realism is neither intuitively correct nor intuitively incorrect, and neither it nor its negation (Weak Conventionalism) is supported by uncontroversial theory, so suspension of judgment about Strong Realism is prima facie justified. Moreover, adopting Ballarin's two schemes together would involve irrationally affirming four jointly inconsistent theses, of which two are logical contradictories (Strong Combinatorialism, Weak Essentialism) and two are logical contraries (Strong Conventionalism, Strong Realism).

Fortunately, other schemes are available. Counting only schemes that assign determinate attitudes to each of the relevant theses, there are 6,561 distinct complete schemes.[8] Granted, most are irrational schemes, not properly available for adoption. Still there are thirty-six distinct rational and complete schemes, on one stringent way of counting rational schemes.[9] We may choose among those thirty-six properly available schemes, for example, on the basis of justificatory considerations.

On a reasonable understanding of *Naming and Necessity*, Kripke does not adopt either of Ballarin's schemes. Kripke rather adopts the following rational and complete scheme: Strong Combinatorialism, denial; Weak Combinatorialism, affirmation; Strong Essentialism, denial; Weak Essentialism, affirmation; Strong Conventionalism, denial; Weak Conventionalism, suspension of judgment; Strong Realism, suspension of judgment; Weak Realism, affirmation. Kripke's scheme is preferable to Ballarin's schemes, at least provisionally, since its assignment of

---

[8] $3^8 = 6{,}561$. This is a liberal estimate of the number of distinct complete schemes, each of which is such that it is psychologically possible to adopt it.

[9] I suppose that a conceptual scheme is *stringently rational* iff it affirms all classical logical truths, denies all classical logical falsehoods, and obeys Łukasiewicz's 3-valued matrices, interpreted so that 0 = denial, 1 = affirmation, and the third value = suspension of judgment. (Using Kleene's matrices instead would make only a small difference here.) So, every stringently rational and complete conceptual scheme on the eight relevant theses affirms, for example, that Strong Combinatorialism is equivalent to the negation of Weak Essentialism, affirms Strong Combinatorialism only if it denies Weak Essentialism, suspends judgment about Strong Combinatorialism if it suspends judgment about Weak Essentialism, and so on.

BOOK REVIEWS

Mild Contraction: Evaluating Loss of Informational Value Due to Loss of Belief. ISAAC LEVI. New York: Oxford University Press, 2004. xi, 245p. Cloth $70.00.

Isaac Levi has frequently summarized the theory of inquiry which he has developed over the years in the formula: infallibility + corrigibility. Agents do not question their beliefs while they act on them; in this sense, their beliefs are infallible for them. But agents realize that their beliefs can undergo changes; in this sense, they must view their beliefs as corrigible. The combination of infallibility with corrigibility has an immediate consequence, already pointed out by Peirce, which much of contemporary epistemology yet seems studiously to ignore: There is no point for an agent to try to justify beliefs that he does not question. What needs justification, however, is giving up beliefs one presently holds to be infallibly true and, perhaps, replacing them by contrary beliefs. On this view, the object of justification cannot be belief but must be belief change.

The book under review is a perfect illustration of this procedure. Levi puts his current views without hesitation or reservation. Yet these views resulted from correcting arguments and theses contained in some of his earlier writings.[1] Thus Levi's attempts to justify his changes of belief about belief change occupy a good part of the book.

But this is not the only aim of the book. In the first chapters Levi sets out clearly and succinctly his views about the aims and proper procedures for inquiry. Given this theory in the background, Levi advocates mild contraction as a principled choice among a number of candidate operations for giving up beliefs while minimizing loss of valuable information. Hence the other aim, and perhaps even the main point of the book, is to exhibit the right reasons for favoring mild contractions and to argue why other reasons that have been offered—by himself and by others—are either insufficient, or misleading, or plain wrong.

Formally, the hero of the book, $\div$, can be quickly introduced as follows. Let K be a belief set (a collection of beliefs closed under logical consequence), let A be a belief, and let—be a contraction

---

[1] See Levi, *The Fixation of Belief and Its Undoing* (New York: Cambridge, 1991), and *For the Sake of the Argument* (New York: Cambridge, 1996).

0022-362X/06/0311/587–91

operation satisfying the AGM-postulates (Alchourrón-Gärdenfors-Makinson). Then define

$$K \div A := \bigcap\{K - A \wedge B : B \text{ is an arbitrary (removable) belief}\},$$

In the principal case that A is not a logical truth; otherwise let $K \div A := K - A = K$. In words: $K \div A$ is what remains as common ground, whenever a choice has to be made between AGM-removing A or some other belief. (The qualification that the belief B be "removable" is not necessary on Hans Rott's account (see below); it is, however, important in Levi's scheme of things: If B is not relevant to the purpose of the contraction at hand, its retraction should not be considered.)

The operation was "found" independently of each other around 1995 by Levi, Maurice Pagnucco and Rott (PR).[2] Disagreement begins over terminology. PR call this operation *severe withdrawal.* Levi prefers the term *mild contraction.* Why? PR simply follow David Makinson's recommendation to refer to subtracting belief change operations that do not satisfy the so-called recovery postulate (such as $\div$ as defined above) as "withdrawal operations." (According to Recovery, the original belief set can be restored after contraction by adding the retracted belief and closing under logical consequence.) Since $\div$ can be shown to be inclusion-minimal with respect to a certain important property (that is, revision equivalence), $\div$ is not mild but severe. PR's terminological choice is thus based on salient formal properties. Levi fears that adopting Makinson's recommendation may lead one to jump to the conclusion that withdrawals are second rate to contractions. If so, Recovery—a principle which Levi rejects—would be canonified by the sheer magic of a word. For this reason Levi prefers calling $\div$ by the good old word "contraction"; the operation is not severe but only mild, because it minimizes loss of valuable information according to Levi. The terminology of the book under review will be adopted in what follows.

Disagreement continues over the best way of presenting mild contractions. The above definition is certainly convenient for those who are familiar with AGM-contraction. For PR, however, mild contraction comes out most naturally, when represented in the so-called spheres models of belief change. Let a belief set K be represented by a set of possible worlds: those worlds deemed possible according to K. If a belief is retracted from K, more worlds must be considered possible. But some possibilities are more remote (less plausible) according to K than others. So let us think of K as being surrounded by spheres of plausibility. (The spheres are nested by set-inclusion and centered on the representation $[K]$ of K—graphically this looks like a horizontally cut onion.) When retracting A from K, the Principle of Informational Economy demands that we should not seriously consider very remote ways in which A may be false. We should go no further than to the closest-to-$[K]$ sphere that contains $\neg A$-worlds. Let the function $cl(\neg A)$ (depending on a given system of spheres around $[K]$) single out that sphere. Take all the $\neg A$-worlds in that sphere and add them to $[K]$. The result, $[K] \cup (cl(\neg A) \cap [\neg A])$, is a representation of $K - A$, the AGM-contraction of K by A.

---

[2] See Rott and Pagnucco, "Severe Withdrawal and Recovery," *Journal of Philosophical Logic,* XXVIII (1999): 501–47.

This sounds natural but there are other options. The Principle of Informational Economy, *strictly understood*, demands that for the purpose of removing A from K, we should admit no more than a single possibility of A being false. Anything else leads to a loss of information that is not strictly required by the purpose of contracting. This option, adding only a single $\neg A$-world to $[K]$, is called *maxichoice contraction* in the literature. Maxichoice-contractions have unattractive properties but the main point here is that the model simply affords no information that would always distinguish a single world from $cl(\neg A)$.

If choosing a single world from $cl(\neg A)$ is no option, what then is so special about the $\neg A$-worlds in $cl(\neg A)$? To be sure, we need to include one such world in order to free us from commitment to A. But many A-worlds in $cl(\neg A)$ are at least as plausible as some $\neg A$-worlds in $cl(\neg A)$. So, if the former be included, should the latter not be included just as well? An affirmative answer is given by the Principle of Weak Preference: If one world is considered as least as plausible as another, then the former should be admitted whenever the latter is. On combining the purpose of contracting with the two Principles, the following operation seems most natural:

$$[K \div A] := cl(\neg A):$$

This is PR's preferred way of presenting mild contractions. It is a natural outcome of combining within the framework of spheres models Informational Economy with Weak Preference.

Levi believes that spheres models "are in the final analysis uninformative in revealing the structure of acceptable contractions" (185). The spheres models afford a comparison of possibilities (with respect to "plausibility") but they do nothing to compare contraction candidates (represented by sets of possibilities) as to the loss of informational value they incur. A serious concern with informational economy, however, requires the latter. While PR allow informational economy to be tempered by other choice principles, Levi insists that informational economy must not be compromised.

The original AGM-modelling in terms of selection functions is afflicted by the same problem. In retracting *A* from *K*, we should want to move to a subset of *K* that no longer entails *A* while keeping as much valuable information from *K* as is compatible with the purpose at hand. In general there are many subsets of *K* that are maximal with respect to the property of not entailing *A*. If informational value is measured by logical strength, we should choose one of these sets as the contraction of *K* by *A* (so-called maxichoice). But it is highly implausible to suppose that choice can always be narrowed down to uniqueness. So AGM recommend a skeptical approach according to which the contraction should be the intersection of those maxichoices that are tied under a selection function. At this point we come to a crucial choice as to the proper philosophical underpinning of the formal theory of belief change. Either we measure informational value by logical strength; in which case we have to say farewell to the Principle of Informational Economy strictly understood. Or we find a measure of informational value according to which the intersection of tied candidates can itself qualify as a candidate by retaining as much informational value as each of the intersected candidates. AGM choose the first, Levi chooses the second option. Without going into the details of Levi's proposal (which, as we have noted, is formally equivalent to PR's notion of severe withdrawal), the ensuing discussion can be summarized thus: Levi criticizes severely those who take the first option. According to Levi these philosophers either do not even pretend to offer a serious philosophical interpretation of their theory or they enter in downright incoherence when they aim at but, according to Levi, fail to deliver a cogent interpretation of contractions in decision-theoretic terms. The prime target of the former type is the AGM-trio while PR serve as the principal target of the latter kind. Levi's verdicts are strong, perhaps even harsh at times. But his perseverance in examining the various proposals in minute detail and in following his arguments to the last consequences is simply admirable.

In the final chapter of the book Levi discusses an argument against Recovery which is based on what has been called the Filtering Condition. The basis of the condition is the observation that some ("generated") beliefs are included in a belief set merely because some other ("generating") belief was included. Filtering requires that once the generating belief is retracted, the generated beliefs should be given up too. Recovery fails under this condition, because it depends crucially on the survival of certain generated beliefs after their generator has been removed. Levi rejects Filtering and the consequent argument against Recovery as being based on the mistaken idea of "pedigree epistemology."

In order to apply Filtering one needs to keep track of some justificatory structure among beliefs. Distinguishing a base from which a belief set is generated by closing under logical consequence is, perhaps, the simplest example of such a justificatory structure. In a sense, such a structure supplies a notion of "pedigree" for beliefs. But contrary to what Levi suggests, the point of Filtering is not to align the theory of belief change with some version of epistemological foundationalism. It is rather to allow a return to a safe state of belief—a "system restore point"—once things have gone wrong.

When things go terribly wrong, that is, when an agent needs to extract himself from inconsistency, Levi himself recommends filtering—though, of course, in different words. Suppose an agent routinely expands a consistent belief set K, containing $\neg A$, by A. The result, $K + A$, is, due to closure under classical consequence, the trivial set $K_\perp$ of all beliefs—epistemic hell" as Levi calls it, following Gärdenfors. It would be nice to return from $K_\perp$ to some sensible belief state by simply contracting $K_\perp$. But since all inadvertent expansions into inconsistency result in the same trivial belief set, all information is lost in order to return to the specific point where the damage had been done.

Levi's solution is simple. Prior to expanding, we should take note of where we are and return to that state, if necessary. In effect (though his wording is different), Levi recommends treating K as a base (in the sense of André Fuhrmann and Sven Ove Hansson) and to consider the consequences ("filter") of $K \cup \{A\}$. If $\perp \in Cn\ (K \cup \{A\})$, then he recommends what has been called a base contraction on $K \cup A$. Since, by hypothesis, A is independent in K, the base contraction amounts to simply deleting A, thus returning to K, and thus removing all the nonsense from the theory $K + A$ which had been there just because A had been added to the base K. As far as recovery from epistemic hell is concerned, filtering seems to be just the right way to undo the original sin.

This book has been written out of a serious and fundamental concern. Levi is a pragmatist and as such he has argued for many years that a theory of justified belief change is a logical continuation of the pragmatist theory of inquiry—as such it receives its philosophical credentials. Now that the topic of belief change has taken a firm place in philosophy and neighboring disciplines, Levi senses the danger that the subject be overgrown by untenable claims and interpretations. The book is an attempt to relate back the theory of belief change to the fundamental purposes of inquiry. It does so by drawing out a broad and important vision in minute detail and with philosophical passion.

ANDRÉ FUHRMANN

São Paulo

NEW BOOKS: ANTHOLOGIES

GARB, BRADLEY P., and J. C. BEALL, eds.: *Deflationary Truth.* Chicago: Carus, 2005. x, 438 p. Paper $33.00.

GARBER, DANIEL, and STEVEN NADLER, eds.: *Oxford Studies in Early Modern Philosophy: Volume II.* New York: Oxford, 2005. x, 259 p. Paper $35.00; *Oxford Studies in Early Modern Philosophy: Volume III.* New York: Oxford, 2006. xi, 246 p. Paper $39.95.

GARCÍA-CARPINTERO, MANUEL, and JOSEP MACIA, eds.: *Two-Dimensional Semantics.* New York: Oxford, 2006. vi, 355 p. Cloth $95.00, paper $35.00.

GARDINER, STEPHEN M., ed.: *Virtue Ethics: Old and New.* Ithaca, NY: Cornell, 2005. 222 p. Cloth $49.95, paper $19.95.

GAUKKROGER, STEPHEN, ed.: *The Blackwell Guide to Descartes' Meditations.* Malden, MA: Blackwell, 2006. ix, 254 p. Cloth $64.95, paper $29.95.

GEHRING, VERNA V., ed.: *Community Matters: Challenges To Civic Engagement In the 21st Century.* Lanham, MD: Roman and Littlefield, 2005. vii, 102 p. Cloth $49.00, paper $17.95; *The Ethical Dimensions of Global Development.* Lanham, MD: Rowman and Littlefield, 2006. vii, 96 p. Cloth $60.00, paper $19.95.

GENNARO, ROCCO J., ed.: *Higher-Order Theories of Consciousness: An Anthology.* Philadelphia: John Benjamins, 2004. xi, 368 p. Cloth $138.00, paper $81.95.

GENDLER, TAMAR SZABÓ, and JOHN HAWTHORNE, eds.: *Oxford Studies in Epistemology, Volume 1.* New York: Oxford, 2005. xi, 340 p. Cloth $74.00, paper $29.75; *Perceptual Experience.* New York: Oxford, 2006. viii, 550 p. Cloth $110.00, paper $35.00.

GEORGE, STEVEN K., ed.: *The Moral Philosophy of John Steinbeck.* Lanham, MD: Rowman and Littlefield, 2005. xv, 201 p. Paper $35.00.

GERSH, STEPHEN, and DERMOT MORAN, eds.: *Eriugena, Berkeley, and the Idealist Tradition.* Notre Dame: University Press, 2006. ix, 318 p. Paper $42.00.

GIBSON, JOHN, and WOLFGANG HUEMER, eds.: *The Literary Wittgenstein.* New York: Routledge, 2004. xi, 356 p. Paper $31.95.

GILL, CHRISTOPHER, ed.: *Virtue, Norms, and Objectivity: Issues in Ancient and Modern Ethics.* New York: Oxford, 2005. xi, 326 p. Cloth $74.00.

GILL, MARY LOUISE, and PIERRE PELLEGRIN, eds.: *A Companion to Ancient Philosophy.* Malden, MA: Blackwell, 2006. xxxvi, 791 p. Cloth $149.95.

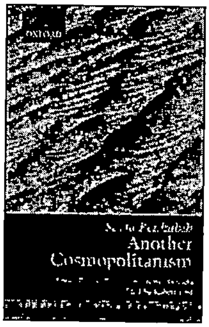
GOEHR, LYDIA, and DANIEL HERWITZ, eds.: *The Don Giovanni Moment.* New York: Columbia, 2006. xxii, 238 p. Cloth $40.00.

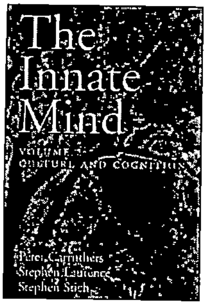
GOODIN, ROBERT E., and PHILIP PETTIT, eds.: *Contemporary Political Philosophy: An Anthology, Second Edition.* Malden, MA: Blackwell, 2006. xiv, 754 p. Paper $49.95.

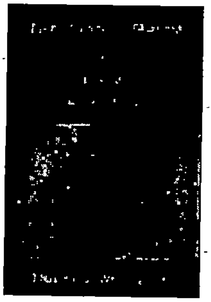
GOODMAN, RUSSELL B., ed.: *Contending with Stanley Cavell.* New York: Oxford, 2005. x, 205 p. Cloth $45.00.

GOLDING, MARTIN P., and WILLIAM A. EDMUNDSON, eds.: *The Blackwell Guide to the Philosophy of Law and Legal Theory.* Malden, MA: Blackwell, 2004. ix, 355 p. Paper $34.95.

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Another Cosmopolitanism
SEYLA BENHABIB
Edited by ROBERT POST

Benhabib argues that since the UN Declaration of Human Rights in 1948, we have entered a phase of global civil society which is governed by cosmopolitan norms of universal justice. The volume contains a substantive introduction by Robert Post, the volume editor, and contributions by panelists Bonnie Honig, Will Kymlicka, and Jeremy Waldron.

October 2006 $19.95

Innate Mind
Volume 2: Culture and Cognition
Edited by PETER CARRUTHERS, STEPHEN LAURENCE, and STEPHEN STICH

"This is a fantastic volume. It beautifully answers the tired complaint that we need to get beyond the simple nature/nurture dichotomy. This collection of essays represents a striking advance in our understanding of both nativism and culture. It also provides the basis for a bright future in charting how innate and cultural components interact."—Shaun Nichols

December 2006 paper $35.00 cloth $85 00

New in Paperback!
The Magic Prism
An Essay in the Philosophy of Language
HOWARD WETTSTEIN

"Wettstein has provided a highly illuminating and thought-provoking anthropology of our practice of using words to refer to things. It provides the perfect antidote to the widespread tendency of distorting this practice because of theoretical fancies and *l'art pour l'art* technicalities."—*Notre Dame Philosophical Reviews*

2004 (paper September 2006) paper $24 95 cloth $55.00

The Oxford Handbook of Philosophy of Language
Edited by ERNEST LEPORE and BARRY C. SMITH

A superb international team contribute forty brand-new essays covering topics from the nature of language to meaning, truth, and reference, and the interfaces of philosophy of language with linguistics, psychology, logic, epistemology, and metaphysics.

November 2006 $160.00

The Architecture of the Imagination
New Essays on Pretence, Possibility, and Fiction
Edited by SHAUN NICHOLS

This volume brings together specially written essays by leading researchers on the propositional imagination. This is the mental capacity we exploit when we imagine that Holmes has a bad habit or that there are zombies. It plays an essential role in philosophical theorizing, engaging with fiction, and indeed in everyday life. *The Architecture of the Imagination* capitalizes on recent attempts to give a cognitive account of this capacity, extending the theoretical picture and exploring the philosophical implications.

November 2006 paper $35.00 cloth $99.00

CAMBRIDGE

BEST IN SCHOLARSHIP

Christ and Horrors
The Coherence of Christology
Marilyn McCord Adams
Current Issues in Theology
$85.00: Hb: 978-0-521-86682-8
$29.99. Pb: 978-0-521-68600-6:
344 pp.

Second Edition!
Nietzsche: *On the Genealogy of Morality* and Other Writings
Student Edition
Edited by Keith Ansell-Pearson
Translated by Carol Diethe
Cambridge Texts in the History of Political Thought
$50.00: Hb: 978-0-521-87123-5
$16.99: Pb: 978-0-521-69163-5:
242 pp.

An Introduction to Political Philosophy
Colin Bird
Cambridge Introductions to Philosophy
$80.00: Hb: 978-0-521-83625-8
$24.99: Pb: 978-0-521-54482-5:
322 pp.

Kant's *Critique of Pure Reason*
An Introduction
Jill Vance Buroker
Cambridge Introductions to Key Philosophical Texts
$75.00: Hb: 978-0-521-85315-6
$24.99: Pb: 978-0-521-61825-0:
336 pp.

Political Theory and the Ecological Challenge
Edited by Andrew Dobson and Robyn Eckersley
$75.00: Hb: 978-0-521-83810-8
$29.99: Pb: 978-0-521-54698-0:
270 pp.

Prices subject to change.

The Cambridge Companion to Hayek
Edited by Edward Feser
Cambridge Companions to Philosophy
$75.00: Hb: 978-0-521-84977-7
$29.99: Pb: 978-0-521-61501-3:
360 pp.

Modal Logic for Philosophers
James W. Garson
$80.00: Hb: 978-0-521-86367-4
$34.99· Pb: 978-0-521-68229-9:
472 pp.

The Worlds of Renaissance Melancholy
Robert Burton in Context
Angus Gowland
Ideas in Context
$90.00: Hb: 978-0-521-86768-9:
356 pp.

A History of Natural Philosophy
From the Ancient World to the Nineteenth Century
Edward Grant
$70.00: Hb: 978-0-521-86931-7
$24.99: Pb: 978-0-521-68957-1:
376 pp.

The Intrinsic Worth of Persons
Contractarianism in Moral and Political Philosophy
Jean Hampton
Edited by David Farnham
$70.00: Hb: 978-0-521-85686-7
$26.99: Pb: 978-0-521-67325-9:
240 pp.

Classical Theory in International Relations
Edited by Beate Jahn
Cambridge Studies in International Relations
$80.00: Hb: 978-0-521-86685-9
$32.99: Pb: 978-0-521-68602-0:
334 pp.

Wordsworth's Philosophic Song
Simon Jarvis
Cambridge Studies in Romanticism
$85.00: Hb: 978-0-521-86268-4:
284 pp.

Oliver Wendell Holmes, Jr., Legal Theory, and Judicial Restraint
Frederic R. Kellogg
$70.00: Hb: 978-0-521-86650-7:
208 pp.

Aquinas, Aristotle, and the Promise of the Common Good
Mary M. Keys
$70.00: Hb: 978-0-521-86473-2:
270 pp.

Principles of Constitutional Design
Donald S. Lutz
$80.00: Hb: 978-0-521-86168-7:
278 pp.

The Cambridge Companion to Atheism
Edited by Michael Martin
Cambridge Companions to Philosophy
$75.00: Hb: 978-0-521-84270-9
$27.99: Pb: 978-0-521-60367-6:
352 pp.

www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

FROM CAMBRIDGE

War Crimes and Just War

Larry May

$80.00: Hb: 978-0-521-87114-3
$29.99: Pb: 978-0-521-69153-6:
360 pp.

Leibniz's Metaphysics

Its Origins and Development

Christia Mercer

$110.00. Hb: 978-0-521-40301-6
$55.00: Pb: 978-0-521-02992-6:
542 pp.

An Introduction to the Philosophy of Language

Michael Morris

Cambridge Introductions to Philosophy

$75.00: Hb: 978-0-521-84215-0
$29.99: Pb: 978-0-521-60311-9:
336 pp.

Arguing about Gods

Graham Oppy

$90.00: Hb: 978-0-521-86386-5:
472 pp.

Gandhi's Philosophy and the Quest for Harmony

Anthony J. Parel

$75.00: Hb: 978-0-521-86715-3:
240 pp.

Toward a Theory of Human Rights

Religion, Law, Courts

Michael J. Perry

$70.00: Hb: 978-0-521-86551-7:
272 pp.

Socratic Virtue

Making the Best of the Neither-Good-Nor-Bad

Naomi Reshotko

$85.00: Hb: 978-0-521-84618-9:
218 pp.

A Natural History of Pragmatism

The Fact of Feeling from Jonathan Edwards to Gertrude Stein

Joan Richardson

Cambridge Studies in American Literature and Culture

$85.00: Hb: 978-0-521-83748-4
$28.99: Pb: 978-0-521-69450-6:
346 pp.

4 Volume Set!

Richard Rorty: *Philosophical Papers*

$120.00: 4 Pb books:
978-0-521-70152-5

Philosophy as Cultural Politics

Philosophical Papers
Volume 4

Richard Rorty

$80.00: Hb: 978-0-521-87544-8
$22.99: Pb: 978-0-521-69835-1:
218 pp.

The Cambridge Companion to Early Modern Philosophy

Edited by Donald Rutherford

Cambridge Companions to Philosophy

$75.00: Hb: 978-0-521-82242-8
$29.99: Pb: 978-0-521-52962-4:
438 pp.

Human Goodness

Pragmatic Variations on Platonic Themes

Paul Schollmeier

$80.00: Hb: 978-0-521-86384-1:
320 pp.

The Heart of Judgment

Practical Wisdom, Neuroscience, and Narrative

Leslie Paul Thiele

$80.00: Hb: 978-0-521-86444-2:
334 pp.

Moral Repair

Reconstructing Moral Relations after Wrongdoing

Margaret Urban Walker

$70.00: Hb: 978-0-521-81088-3
$27.99: Pb: 978-0-521-00925-6:
262 pp.

Fundamentals of Critical Argumentation

Douglas Walton

Critical Reasoning and Argumentation

$75.00: Hb: 978-0-521-82319-7
$29.99: Pb: 978-0-521-53020-0:
360 pp.

The Cambridge Companion to Tocqueville

Edited by Cheryl B. Welch

Cambridge Companions to Philosophy

$80.00: Hb: 978-0-521-84064-4
$29.99: Pb: 978-0-521-54996-7:
460 pp.

Natural Law Liberalism

Christopher Wolfe

$75.00: Hb: 978-0-521-84278-5:
268 pp.

Prices subject to change

www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

New books from The MIT Press

Hot Thought

Mechanisms and Applications of Emotional Cognition

Paul Thagard

"Impressively comprehensive, unfailingly sensible, and made all the more appealing by its hip-pocket readability, *Hot Thought* will be a godsend to instructors in philosophy and cognitive science."
— Patricia S. Churchland, University of California, San Diego

A Bradford Book • 320 pp., 34 illus. $36

De Lingua Belief

Robert Fiengo and Robert May

"Takes on some of the hardest and longest-standing problems in the philosophy of language." — Michael Glanzberg, University of California, Davis

A Bradford Book • 224 pp. $32

Relativism and the Foundations of Philosophy

Steven D. Hales

"An original and stimulating book whose conclusions will challenge deep-seated orthodoxies and provoke strong reactions." — Maria Baghramian, University College Dublin

A Bradford Book • 256 pp. $36

Body Language

Representation in Action

Mark Rowlands

"Here is one of those rare books that might actually change the way philosophers and cognitive scientists think."
— Andy Clark, University of Edinburgh

A Bradford Book • 218 pp. $36

Persistence

Contemporary Readings

edited by Sally Haslanger and Roxanne Marie Kurtz

Influential accounts of persistence provide an overview of the topic and explore in depth both the pros and cons of the perdurantist, exdurantist, and endurantist answers to questions about how ordinary objects persist through time.

MIT Readers in Contemporary Philosophy
A Bradford Book • 496 pp., 26 illus. $36 paper

Heidegger's Topology

Being, Place, World

Jeff Malpas

"A brilliant book that will change the entire field of Heidegger studies."
— Edward S. Casey, Stony Brook University

456 pp. $38

Critique and Disclosure

Critical Theory between Past and Future

Nikolas Kompridis

"A bold and innovative book that attempts to set out an agenda for critical theory that reaches beyond Habermas."
— James Bohman, St. Louis University

368 pp. $37.50

Consciousness, Function, and Representation

Collected Papers, Volume 1

Ned Block

The first of a planned two-volume collection of Ned Block's writings on philosophy of mind; this volume treats consciousness, functionalism, and representation and can be regarded as Block's most complete statement of his positions on consciousness.

A Bradford Book • 680 pp., 16 illus. $34 paper

Rediscovering Empathy

Agency, Folk Psychology, and the Human Sciences

Karsten R. Stueber

This is the clearest and most compelling interdisciplinary analysis of empathy available, essential reading for philosophers, psychologists, and social scientists."
— Bertram F. Malle, University of Oregon

A Bradford Book • 288 pp. $36

Rationality and Logic

Robert Hanna

"This stimulating and wide-ranging book will be of interest to philosophers of logic and also to cognitive psychologists."
— Professor Jane Heal, University of Cambridge

A Bradford Book • 336 pp. $35

Dark Ages

The Case for a Science of Human Behavior

Lee McIntyre

"Reminds us that we have a moral and intellectual obligation to seek the fullest possible understanding of the roots of human behavior. McIntyre has written a beautiful and timely ode to scientific rationality." — Sam Harris, author of *The End of Faith* and *Letter to a Christian Nation*

192 pp. $24.95

Conceptual Issues in Evolutionary Biology

Third Edition

edited by Elliott Sober

Essays by philosophers and scientists address conceptual issues in evolutionary biology.

A Bradford Book • 632 pp., 26 illus. $38 paper

now in paperback

Brain Fiction

Self-Deception and the Riddle of Confabulation

William Hirstein

"An excellent starting point for thinking about the biological mechanisms that help us make sense of the world." — *American Scientist* online

Philosophical Pathology: Disorders in Mind series
A Bradford Book • 288 pp., 16 illus. $18 paper

now in paperback

Varieties of Meaning

The 2002 Jean Nicod Lectures

Ruth Garrett Millikan

"Not only linguists and cognitive psychologists but also philosophers interested in mind and language (two central topics in contemporary analytic philosophy) will find Millikan's book very appealing." — *Choice*

Jean Nicod Lectures • A Bradford Book
256 pp., 3 illus. $18 paper

now in paperback

Imagination and the Meaningful Brain

Arnold H. Modell

"A profound rumination on the perennial problem of how to understand other minds." — *Journal of Consciousness Studies*

A Bradford Book • 272 pp., 5 illus. $18 paper

A Theory of General Ethics

Human Relationships, Nature, and the Built Environment

Warwick Fox

"A very ambitious, very original attempt to provide ethical principles of human relations with the natural, the social, and the humanly built worlds. It makes a serious and very important contribution to philosophy." — Roger S. Gottlieb, author of *A Greener Faith*

400 pp. $28 paper

The Act of Being

The Philosophy of Revelation in Mulla Sadra

Christian Jambet

translated by Jeff Fort

Exploring the thought of Mulla Sadra Shirazi, an Iranian Shi'ite of the seventeenth century: a universe of politics, morality, liberty, and order that is indispensable to our understanding of Islamic thought and spirituality.

Distributed for Zone Books • 424 pp. $38.95

now in paperback

How the Mind Explains Behavior

Folk Explanations, Meaning, and Social Interaction

Bertram F. Malle

"This is a significant contribution to psychology, and it will have lasting value." — Bernard Weiner, University of California, Los Angeles

A Bradford Book • 328 pp., 20 illus. $19 paper

now in paperback

Sweet Dreams

Philosophical Obstacles to a Science of Consciousness

Daniel C. Dennett

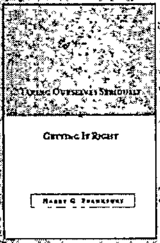
"In characteristically playful mode, Darwinian fundamentalist Daniel Dennett turns his opponents' arguments against them in a masterful display of philosophical judo." — *Toronto Globe & Mail*, The Globe 100, the "Best and Most Influential Books of 2005"

A Bradford Book • 216 pp., 3 illus. $14.95 paper

To order call **800-405-1619.**

http://mitpress.mit.edu

NEW FROM STANFORD UNIVERSITY PRESS

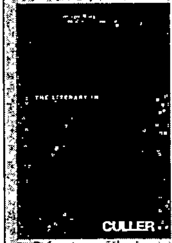


Taking Ourselves Seriously and Getting It Right

HARRY G. FRANKFURT

With the same insight and humor he brought to *On Bullshit*, Harry G. Frankfurt, in this volume of his Tanner Lectures in Moral Philosophy, tackles the provocative question, "What is it about human beings that makes it possible for us to take ourselves seriously?"

$14 95 paper



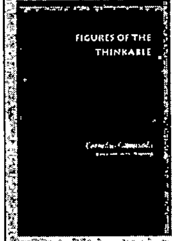
The Literary in Theory

JONATHAN CULLER

"*The Literary in Theory* takes up questions that have been basic to the enterprise known as 'theory' with a fine mix of historical awareness, lively critical sense, and thoughtful advocacy. . . . Is theory dead? This book shows that it has a pulse and a sense of humor."

—Haun Saussy,
Yale University

$21.95 paper $55.00 cloth

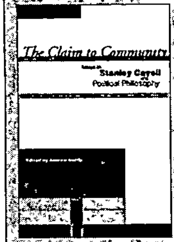


Figures of the Thinkable

CORNELIUS CASTORIADIS, Translated by HELEN ARNOLD

In this posthumous collection of articles, lectures, and interviews Cornelius Castoriadis pursues his incisive analysis of modern society, the philosophical basis of our ability to change it, and the points of intersection between his many approaches to this theme.

$24.95 paper $65.00 cloth



The Claim to Community

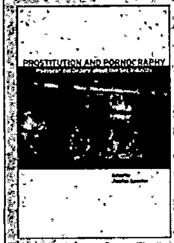
Essays on Stanley Cavell and Political Philosophy

Edited by ANDREW NORRIS

"This marvelous collection of essays is the best available examination of the most profound and interdisciplinary American philosopher writing today: Stanley Cavell."

—Cornel West,
Princeton University

$24.95 paper $65.00 cloth



Prostitution and Pornography

Philosophical Debate About the Sex Industry

Edited by JESSICA SPECTOR

Bringing together academics and industry veterans, this collection of new and classic writings about the sex industry asks us to think about the differences between society's treatment of prostitution and pornography, while investigating how liberalism deals with the sex industry in general.

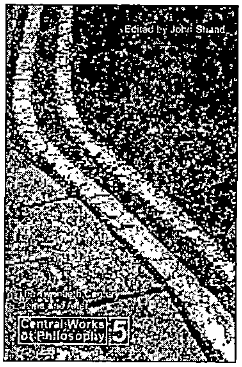
$27.95 paper $70.00 cloth

Stanford

800.621.2736 www.sup.org University Press

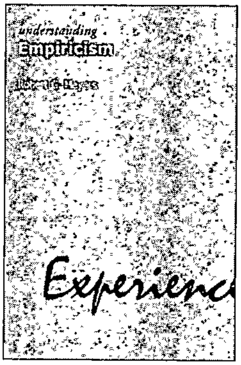
Central Works of Philosophy, Volume 5
The Twentieth Century: Quine and After
John Shand
Central Works of Philosophy 5
0-7735-3083-5 $24.95 US / $29.95 CA paper
0-7735-3082-7 $85.00 cloth

VOLUME 5 covers the central texts in the history of analytic philosophy, from Quine's *Word and Object* to McDowell's *Mind and World*. The texts range over political philosophy, ethics, metaphysics, and the philosophies of language, mind, and logic. Representing some of the most important philosophical work of the last forty years, these essays provide a map and compass for the current philosophical landscape.



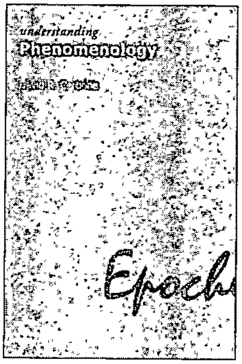
Understanding Empiricism
Robert G. Meyers
Understanding Movements in Modern Thought 5
Acumen Publishing
1-84465-059-6 $22.95 US / $27.95 CA paper
1-84465-058-8 $85.00 cloth

"A welcome addition to the introductory literature on empiricism, filling a gap between the books which are very heavy on Locke, Berkeley, Hume scholarship, and books that focus purely on contemporary epistemology"
Graham Stevens, University of Manchester



Understanding Phenomenology
David R. Cerbone
Understanding Movements in Modern Thought 4
Acumen Publishing
1-84465-055-3 $22.95 US / $27.95 CA paper
1-84465-054-5 $85.00 cloth

"Essential reading for undergraduate classes in phenomenology." Sean Kelly, Princeton University



McGILL
QUEEN'S
UNIVERSITY
PRESS
www.mqup.ca

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details.